A Computational Theory of Goal-Directed Style in Syntax

Chrysanne DiMarco*
University of Waterloo

Graeme Hirst†
University of Toronto

The problem of style is highly relevant to computational linguistics, but current systems deal only superficially, if at all, with subtle but significant nuances of language. Expressive effects, together with their associated meaning, contained in the style of a text are lost to analysis and absent from generation.

We have developed an approach to the computational treatment of style that is intended to eventually incorporate three selected components—lexical, syntactic, and semantic. In this paper, we concentrate on certain aspects of syntactic style. We have designed and implemented a computational theory of goal-directed stylistics that can be used in various applications, including machine translation, second-language instruction, and natural language generation.

We have constructed a vocabulary of style that contains both primitive and abstract elements of style. The primitive elements describe the stylistic effects of individual sentence components. These elements are combined into patterns that are described by a stylistic meta-language, the abstract elements, that define the concordant and discordant stylistic effects common to a group of sentences. Higher-level patterns are built from the abstract elements and associated with specific stylistic goals, such as clarity or concreteness. Thus, we have defined rules for a syntactic stylistic grammar at three interrelated levels of description: primitive elements, abstract elements, and stylistic goals. Grammars for both English and French have been constructed, using the same vocabulary and the same development methodology. Parsers that implement these grammars have also been built.

The stylistic grammars codify aspects of language that were previously defined only descriptively. The theory is being applied to various problems in which the form of an utterance conveys an essential part of meaning and so must be precisely represented and understood.

1. Introduction

1.1 An Advocacy of Style

Understanding a text requires more than just understanding its propositional content. It requires a sensitivity to the interaction of semantic content, emotional expression, and interpersonal and situational attitudes. This interaction is reflected in the *style* of the text. Style in language is not just surface appearance, a decorative veneer. Rather, it is an essential part of meaning, part of the author's communication to the reader. So to fully understand the nuances of a text, one must determine not only the propositional content, but also how its communicative effect is colored by the form, which reflects affective content. While propositional content provides the basic *tone*, the expressive

---

* Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.
† Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4.

form provides the *tonal quality*. Together, form and content create style, that which distinguishes both an individual text and a collective body of writing.

Style is created through subtle variation, seemingly minor modulations of exactly what is said, the words used to say it, and the syntactic constructions employed, but the resulting effect on communication can be striking. Consider the following versions of the same text, Matthew 7:27:

1. And descended the storm and came the floods and blew the winds and beat against that house and it fell and the fall of it was great. (Literal translation of the Hellenistic Greek.)

2. And the rain descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it. (Revised Standard Version, 1952.)

3. The rain came down, the floods rose, the wind blew, and beat upon that house; down it fell with a great crash. (The New English Bible, 1970.)

4. Rain came down, floods rose, gales blew and struck that house, and it fell; and what a fall it had! (The New Jerusalem Bible, 1985.)

5. The rains fell, the torrents came, the winds blew and lashed against his house. It collapsed under all this and was completely ruined. (The New American Bible, 1973.)

The first variation, a word-by-word translation from the Hellenistic Greek, is actually a good deal less striking than subsequent, widely accepted versions. The placement of the verbs before their subjects is quite normal and the closing is also quite usual. The sense of discord and resulting poetic effect is not as evident as it is in the second variation, from the Revised Standard Version, which is resoundingly poetic in the imitative form of the first five clauses, followed by the inverted form of the final clause. The text begins in strong concord and dissolves into discord, but not unpleasingly so. *Discord*, as we use the term in this paper, refers to a deviation from the norm, but such deviations can be used to good effect. As we will see, in language as in music, it is often through the construction of patterns of concord *and* discord, particular combinations of order and disorder that create an overall harmonious arrangement, that certain stylistic effects are achieved.

In versions 3 and 4, the dramatic effect of the closing has been retained from the RSV translation but through different choices of words and structure. In 3, the last clause is inverted, but the translator has chosen to place more emphasis upon the fall itself (*down it fell*) than its magnitude (*great*). Example 4 ends on an intense note, achieved through the use of an exclamation (*what a fall it had!*), rather than an inversion of syntactic structure. And no single word expresses the degree of magnitude of the fall. In the final example, 5, the drama and intensity of the RSV has been lost with the removal of the dissolution from initial concord to final discord. In this case, the translator has opted for plainness and clarity, even at the expense of beauty.

In English, example 1 sounds odd and disjointed; and examples 2, 3, and 4 are dramatic in a way that example 5 is not. But all have the same essential content. What then causes the differences in effect? What is being varied? There are at least four parameters that play a role in these stylistic variations: *lexical, syntactic, thematic*, and *semantic* aspects.

Lexical aspects: Compare example 4 to the following constructed versions:[1]

    6.   Rain descended, floods rose, gales raged and beat upon that dwelling, and it collapsed; and what a fall it had!

    7.   Rain fell, the water level rose, winds blew and hit that house, and it fell; and what a fall it had!

The differences between 4, 6, and 7 are primarily lexical. Example 6 uses rather elegant words, while 7 opts for a more commonplace vocabulary, and 4 lies somewhere between the two in lexical formality.

Syntax: Comparing example 2 to example 5, we see that the former uses syntactic structures that create more dramatic effects: a short climactic sentence, *and it fell*, and a striking inversion, *great was the fall of it*, close the text on a powerful note. In contrast, 5 ends with a very ordinary, unremarkable, construction.

Thematic aspects: Comparing example 2 to example 3, we observe that the two sentence structures bring different elements into focus: *great was the fall of it*, in contrast to *down it fell with a great crash*. In general, variations in thematic structure can create different stylistic effects.

Semantics: The exact choice of what is said, or not said, also has stylistic consequences. Compare example 4 to the following two constructed versions:

    8.   A serious storm, with rain and gales and floods, struck that house, which collapsed.

    9.   Rain fell and fell and eventually caused a flood, which rose up to that house; also the winds kept blowing until eventually the combined forces of rain, flood, and wind were too great and caused so much structural damage that the house collapsed.

These variations differ in their semantic content: example 4 merely reports the events, but 8 evaluates, *a serious storm*, while 9 emphasizes technical details, *structural damage*. These semantic differences are reflected in the texts; and stylistic differences result.

    Given that these four parameters—lexical choice, syntax, theme, and semantics—control stylistic variations, two questions arise:

- How do we characterize each type of variation?

- How does each variation contribute to an overall stylistic effect?

In considering these questions in this paper, we will concentrate on variations of the *syntactic* parameter.

    We emphasize that our intent here is *not* any form of literary analysis or literary theory. Our main concern will be ordinary, everyday text. It is true that, in the Biblical texts above, we saw how different syntactic forms carry different stylistic import to the point that one form may be poetry while another is just dull, plodding prose. But style

---

1 We thank Eduard Hovy for constructing examples 6 through 9.

isn't just a matter of achieving poetry or not. Every text, large or small, interesting or dull, effective or not, has its own style. Nor are we talking about style in any of its *normative* senses: the common tenets dictating standard forms (e.g., Chicago 1982), or the textbook prescriptions for 'good' style: be clear, be simple, be precise (e.g., Strunk and White 1979). For people aim to convince, to persuade, to impress, and, even, sometimes to obscure, and standard textbooks tell us very little about such varied and subtle stylistic and pragmatic goals.

Rather than a study of literary or normative style, our intent is to determine what gives any ordinary piece of text its stylistic 'feel.' The following examples will demonstrate what we have in mind. These texts are all from newspaper feature articles:

10. Silvia, a commanding woman in her 50's, a shrew falsely mellowed by religion, promptly organized prayer sessions on the lines of Tupperware meetings.[2]

11. The artist provides a dreamy background done in yellow and bistre brushstrokes to a blue gown with woodenly rigid folds or the profile of a brown angel painted so mineral hard and modeled so carefully that the incoherence of virtue does it injury.[3]

12. Crazed with fear, he tried to purify her by dunking her in the ocean and holding her under the water; then in desperation he threw her on the still-smoking pyre.[4] ᵢ

In a newspaper, we might have expected the writer and translator to have simply aimed for clarity. In fact, we find a variety of effects. The first text, 10, emphasizes a sense of harmony by repeating the same kind of structure, a nominal group, in the postmodification of *Silvia*. The second, 11, is more complex, and achieves a certain balance in the judicious use of conjunctions; but the result is so difficult to understand that it doesn't really make sense. Text 12 has a stark initial participle clause, *crazed with fear*, that emphasizes the intensity of the subject's emotional state.

To account for the kinds of complex stylistic effects that occur even in everyday writing, we propose a *goal-directed* understanding of style. That is, an author's intent can vary with respect to a number of stylistic goals, such as clarity or obscurity, abstraction or concreteness, staticness or dynamism. Particular choices of words, syntactic structure, and semantic structure make a text more—or less—stylistically varied and effective. We believe that these choices, goals such as abstraction or concreteness, and the stylistic elements that are used to realize them can be recognized and represented in a formal notation. They are, in a word, *codifiable*. It is this codification that is at the heart of *computational stylistics*.[5]

1.2 The Function of Style
Propositional content alone is insufficient to determine the nature and form of a sentence (Halliday 1985; McDonald and Pustejovsky 1985; Jameson 1987; Hovy 1988; Scott

---

2 Adapted from The *Manchester Guardian Weekly*, 7 February 1988. Translated from *Le Monde*.
3 The *Manchester Guardian Weekly*, 14 February 1988. Translated from *Le Monde*. .
4 The *Manchester Guardian Weekly*, 7 February 1988, with minor punctuation corrections. Translated from *Le Monde*.
5 The term *computational stylistics* has been used by Milic (1982) to describe what might better be called computer-*aided* stylistics, in which computer-generated data and pattern-matching aid human analysis and judgment of style in literary studies (see Section 2.2). In contrast, our use of the term entails fully automatic computer analysis of the style of any kind of text.

and de Souza 1990). Even after the propositional content has been decided upon, there are still many linguistic, syntactic, and even semantic decisions that the language producer must make before a sentence can be formed. These decisions are assumed by the audience not to have been made randomly, but rather in specific, deliberate ways that encode additional information, such as opinion, emotional affect, and interpersonal relationships.[6] To the extent that a piece of text exhibits a particular, recognizable style, it also reflects the author's presumed intent to convey the effect associated with that style. Therefore, full understanding of a text must represent not only propositional content, but also stylistic effects.

This becomes most apparent in the case of machine translation. If a translation is to be faithful, the stylistic effects of the source language text must be transferred to the target language text, making appropriate use of the stylistic conventions of the target language. But a dilemma arises:

- One wants to *preserve* the original author's stylistic intent, the information being conveyed through the manner of presentation. However, different languages might realize this effect in different ways. So the source and target texts should both aim for the author's stylistic goal, but might have to achieve it through different linguistic means.[7]

- Yet one wants to produce a text whose style is appropriate and natural to the particular target language. Languages differ as to the most 'natural' way to express an idea. For example, French tends to prefer more abstraction, English more concreteness (Vinay and Darbelnet 1958). The best translation, therefore, might *modify* the original author's stylistic intent and express a different effect.

Sometimes, there is no way to resolve this dilemma, and one is left with an unsatisfactory translation. But, with a knowledge of the *comparative stylistics* of a language pair, and of the stylistic resources of each language and the possible range of effects they can create, one can substantially improve the quality of a translation.

1.3 The Structure of the Paper

Our goal is to create a formal representation of stylistics for use in natural language systems, and, moreover, to do so in a manner applicable to different languages. The solution we will propose is a codification of syntactic stylistic knowledge in the form of a *stylistic grammar*.

In the next section, we will review the current status of the codification of style. In Section 3, we will construct a *vocabulary* for stylistics. The definition of concepts and the attempt to organize them into a recognizable structure is a necessary first step toward understanding the problem. Then we will develop in Section 4 a *methodology* for converting stylistic knowledge into a formal representation. The methodology will

---

6 Thus the speaker or writer is *accountable* for his or her stylistic decisions, in the sense that that term is used in Ethnomethodology and Conversation Analysis. A person encountering friends on the street, for example, may choose to greet them or not, but is held accountable either way—failure to greet is a snub; it is not possible to opt out of the situation altogether (Heritage 1984). Similarly, a speaker or writer is accountable for all the stylistic nuances of his or her utterances; it is not possible to utter a sentence in such a way that only the propositional content counts and not the form in which it is expressed.

7 "Economy is expressed at two levels, lexical and syntactic, which may, however, be related: what is expressed lexically in one language may be realized through syntactic means in the other, and vice versa" (Vinay and Darbelnet 1958, p. 185, authors' translation).

be as general as possible, ideally applicable to the codification of style for the sentence and paragraph levels; appropriate for lexical, syntactic, and semantic style; and applicable to both English and French. The methodology will be used in conjunction with the stylistic vocabulary to guide the systematic collection and creation of stylistic rules. This grammar will provide a formal description of the syntactic patterns that differentiate the various stylistic goals.

In addition to theoretical problems, there are implementation issues to be addressed as well. In Section 5, we will describe the development of a *stylistic parser* that uses the formal bodies of rules to analyze complex English text.

In the last section, we will review the contributions of the paper and discuss completed and ongoing extensions to the research.

2. Background to the Study of Style

2.1 Views of Style
2.1.1 The Classical View.
The study of style has an ancient history, for the concept first appeared in the fifth century B.C. Scholars of the time conceived the notion that the rhetorical form of a writer's text should reflect his thoughts and intentions. This was to be achieved through the appropriate choice and organization of words and syntactic structures in order to "evoke the desired response" (Corbett 1971). Content and form were recognized as inseparable, and as exerting a reciprocal influence upon each other.

Beginning with Corax of Syracuse, rhetoricians developed methods for systematic instruction in the art of writing. The classical scholars, from Isocrates and Aristotle to Cicero and Quintilian, established standards of rhetoric that influenced curricula for centuries. The contribution of Aristotle was particularly notable. He countered Plato's argument that rhetoric was mere sophistry, deceitful reasoning, by providing a system of instruction that treated the theory and practice of style as a valid discipline, designed to appeal simultaneously to reason and emotion.

This classical approach to style flourished into the Middle Ages, for it formed part of a standard university education. The teaching of formal rhetoric in schools and universities continued to play a significant role throughout the Renaissance and into the 18th century. However, in the 19th century, the teaching of rhetoric gave way to the teaching of composition, which came to be associated merely with a set of basic prescriptive rules. This approach was in direct contrast to the original classical theory that stressed the education of the whole person as a preparation for achieving style in writing (Corbett 1971).

2.1.2 The Modern View.
Today, the common view of style is still one of conformity to standard good taste and an avoidance of bad form. Textbooks (Fowler 1968; Strunk and White 1979; Kane 1983; Grevisse 1986) promote the idea that there is a universal and correct mode of expression. However, normative rules, while necessary in some degree so that communication is possible at all, may, if too inflexible, deaden the expression of a writer's individual voice. Universal rules of good style are best used, we suggest, simply as a guide to avoiding bad constructions and obstacles to clear writing, but *not* as a basis for a theory of stylistic effects in text.

2.1.3 Our View.
In computational applications, where we expect to deal with large amounts of similar types of text, the analysis of *group style* is of more interest than the idiosyncratic style of any one writer. Group style can be subdivided into two major types, each associated with a different view of stylistics: *literary* style and *utilitarian*

(or *functional*) style. Utilitarian group styles are distinguished from the literary type by their association with a genre of text that has a particular function or purpose, such as medical textbooks or political newspaper writing. In such styles, the writer adjusts her language to what readers expect in a specific situation. The fact that functional group styles *are* somewhat more restrictive than literary group styles, as they are characteristic of a particular genre, suggests that the problem of codifying utilitarian style will be more tractable than the literary case. For this reason, we will focus on utilitarian texts in general, drawing most of our examples from newspaper and magazine journalism. (Our main source of examples is the English translations, published in *The Guardian Weekly*,[8] of articles from *Le Monde*.)

2.2 The Current State of Computational Stylistics

Most of the research to date in computational stylistics has been the development of so-called *style-checkers*. The UNIX-based Writer's Workbench, Grammatik I and II, RightWriter, PC-Style, Punctuation and Style, Electric Webster, and CRITIQUE are examples (Raskin 1986). However, none of these systems is appropriate for our purposes. These programs enforce the basic virtues: be clear, be simple, be precise. They merely check for common grammatical errors such as number disagreement, pronoun case problems, unbalanced punctuation, split infinitives, excessive sentence length, and excessive sentence complexity. There is no systematic approach to constructing a vocabulary of style, no structured representation of stylistic rules. Stylistic knowledge may simply consist, as in the case of CRITIQUE, of an unstructured mass of rules (Jensen et al. 1986, p. 190). These programs have no real understanding of the significance of stylistic parameters.

The other main kind of computational research in stylistics has been statistical (e.g., Cluett 1976, 1990; Milic 1982; Biber 1988, 1989). With this approach, a count is made of the number of occurrences in a sample text of a stylistically significant feature, such as an initial prepositional phrase. A statistical analysis is done to compare the count and type of stylistic features with the corresponding parameters of texts written by authors working in the same or different genres. But the interpretation of the statistical results must be done by a human; the computer system has no real understanding of what the variations in style mean.

Although virtually all computational research in stylistics takes a non-AI approach to the representation and use of stylistic knowledge, the usefulness and degree of sophistication of some of these programs should not be underestimated. The York Computer Inventory of Prose Style (Cluett 1976, 1990) has been undergoing development since 1970. The principal contribution of the project is its convincing demonstration that the richness and subtlety of literary style could in fact be correlated, to a perhaps surprising degree of accuracy, with a catalog of syntactic patterns. But the program that performs the stylistic analysis has no understanding of the meaning of its statistical results; it relies on human interpretation of the statistics computed. Nevertheless, the York project's findings are still very useful, as they have catalogued the syntactic features that produce particular stylistic effects.

Thus, most current approaches in computational stylistics have not included formal stylistic rules, which are clearly a desirable prerequisite. We now review the extent to which such rules have been developed in theoretical stylistics and used in computational stylistics.

The title of Vinay and Darbelnet's (1958) book, *Stylistique comparée du français et de*

---

8 Formerly The *Manchester Guardian Weekly*.

l'anglais, is somewhat misleading, for their purpose was not just to deal with comparative stylistics in translation, but also to enumerate the more common rules and procedures in the overall translation process. But they did not construct a specific vocabulary for comparative stylistics and, in fact, virtually all of their 'stylistic' terms refer to basic syntactic constituents. Using their vocabulary of terms, Vinay and Darbelnet defined rules, expressed in ordinary natural language, that encoded knowledge about French–English comparative stylistics. The rules certainly fell short of the repertoire needed by a professional translator, but no one has yet improved upon their codification of comparative stylistics (Brian Fitch, personal communication). Their book is still used as a prescribed text in translation courses. Vinay and Darbelnet did not take an explicitly goal-directed approach, but their book does contain a great deal of information about the correlation of stylistic goals with particular lexical, syntactic, and semantic choices. Unfortunately, however, their rules proved to be too language-dependent for the framework that is proposed in this paper. But the success of a goal-directed codification, albeit an implicit one, supports the approach that we will take below.

Crystal and Davy (1969) hypothesized the creation of a formal stylistic grammar, built upon a vocabulary of stylistic terms and intended to represent the sentence structures characteristic of a particular stylistic sublanguage. However, the grammar that they sketched contained virtually no terms other than purely syntactic ones and no correlations with stylistic goals; it was only a first step, although an ambitious one, toward a codification of the stylistic rules that define a sublanguage. What we have found especially useful to adapt from their work is the overall methodology, the approach to formalizing style, even the belief that style *can* be formalized.

Kane's (1983) rhetoric and handbook of style represents the typical textbook that aims to teach the rules of classic good writing. Kane, too, did not take an explicitly goal-directed approach in his stylistic rules, but nevertheless gave numerous references to the stylistic goals associated with particular syntactic structures. The systematic construction of a stylistic vocabulary seems not to have been an explicit goal, but he used many terms unique to stylistics. However, his terminology was apparently augmented whenever the need arose to describe a new feature; there was no underlying structure, no formal representation of rules. The stylistic rules were expressed in ordinary English and achieved, as the author intended, good coverage of plain writing style. This body of rules was 'structured' to the extent that there were hierarchies of stylistic sentence types, but, in general, Kane's rules appear to be just an enumeration of stylistic tenets, not a deliberately constructed organization. Nevertheless, we have found it useful to draw upon these rules as justification for the syntactic coverage of our grammar.

In computational stylistics, an application area of particular interest to us is machine translation (MT). In suggesting the applicability of computational stylistics to MT, Loffler-Laurian (1986) emphasized an important point: although corresponding group styles may exist across languages, the realization of the style can be different for each language. Tsutsumi (1990) presented a methodology for bridging *stylistic gaps* (stylistic differences) between the syntax of the source and target languages in MT. An important contribution of Tsutsumi's work is the recognition that computational stylistics is useful for the translation of pairs of languages that are not in the same language group.

Hovy's (1988) PAULINE system was the first computational system that implemented goal-directed style. PAULINE was able to generate text that conformed to various stylistic and pragmatic constraints that it was given. The system was goal-directed, able to correlate such stylistic goals as formality, simplicity, and respect with the lexical

and syntactic characteristics of the text produced. Hovy's method was based on the definition of *rhetorical goals of style*, such as formality, force, and partiality, that control broader pragmatic goals. For example, a low degree of formality combined with high degrees of force and partiality gives a 'no-nonsense' effect.

To achieve a particular rhetorical goal, PAULINE examined the options at various points during the production of text and applied the appropriate strategies. For example, to achieve *formality*, the generator would perform the following actions:

Organization of topics: Make long complex sentences by subordinating them in relative clauses or by conjoining two or more sentence topics.

Organization of sentences: Make sentences seem weighty by including a number of adverbial clauses, by placing these clauses toward the beginning of a sentence, by building parallel clauses, by using the passive voice, by using more 'complex' tenses such as the perfect tenses, and by avoiding ellipsis.

Choice of words and phrases: Select formal words and phrases. Avoid doubtful grammar, popular idioms, slang, and contractions.

Although structured in its correlation of the generator's actions with decision points, this approach to representing a goal-directed knowledge of style is essentially heuristic.

Hovy's success in implementing a goal-based notion of style, even though limited in scope and informal in the mode of knowledge representation, encouraged our development of a computational approach to goal-directed style.

2.3 Summary

- From existing research in theoretical stylistics, we have seen that there is a basis for the codification of group-based, utilitarian, goal-directed stylistics.

- The need for a vocabulary of stylistic terms is an accepted idea. However, the nature and structure of such a vocabulary for computational use, which must describe more than basic syntax and be amenable to systematic construction, have not yet been addressed. These issues will be subjects of the next section.

- Similarly, while large bodies of stylistic rules already exist, they have not been organized into the formal structure necessary to a computational approach. Section 4 will develop a structured method for representing stylistic rules by constructing formal grammars.

- The feasibility of a goal-directed analysis of style is supported by previous research, but an actual goal-based codification has not previously been attempted. In the grammar that we will develop in Section 4, we will incorporate such a goal-directed knowledge of stylistics.

3. A Vocabulary of Style

3.1 Stylistic Goals

If we hope to build AI-based systems to deal with matters of style, then we must provide a formal knowledge representation: we need a vocabulary of well-defined, expressive terms that will allow our intuitions about style to be stated precisely and

understood clearly. We will construct a vocabulary from terms that are associated with utilitarian, group-based, goal-directed stylistics. We will start by considering the kinds of stylistic goals to be dealt with in the lexicon.

When a writer composes text, she has certain stylistic goals in mind; such goals might include clarity, or informality, or even the clouding of an unpalatable message. To achieve her various goals, she will choose specific words (*outplacement* or *firing*, for example), syntactic constructions (heavily modified or sparse noun phrases), and semantic organization (*Thirty percent of the class failed* or *The majority of students passed*).

In applications such as machine translation, we want to understand *why* a writer has used language in a particular way, what specific effects she intended to convey, and which linguistic choices were made to achieve these goals. There are a multitude of goals that could be considered; indeed, the space of possibilities has never been fully explored. For this study, we have chosen six goals that Vinay and Darbelnet (1958) consider to be among the most commonly used, representing the opposite ends of three dimensions. The goals are listed as follows, with an intuitive description of the corresponding characteristics that a sentence of each type would display:

- **Clarity:** Simplicity, harmony, and no ambiguity.

  Obscurity: Complexity, incongruity, and difficulty of understanding.

- **Concreteness:** Specificity, with an emphasis on certain components.

  Abstraction: Generality, with no particular emphasis on any one sentence component.

- **Staticness:** Uniformity, predictability, continuity.

  Dynamism: Deviation from the norm, unexpectedness, action.

3.2 From Theoretical Stylistics to Computational Stylistics

We will now develop a vocabulary of basic terms that will describe stylistically significant aspects of syntax. Our development is guided by much in classical rhetorical theory; often, our definitions will be extensions of those of the classical theory. We do this in two stages: first, we outline abstract properties of style suggested by classical rhetoric; second, we convert these properties into definitions of abstract elements that will be used in a grammar of style.

Our starting point is the notion of *stylistic norm*, which is defined as the most commonly used structures in a given genre. This notion is intrinsic to the work of the majority of descriptive stylists such as Crystal and Davy (1969) and Cluett (1976, 1990). From this, we derive the notions of stylistic *concord* and *discord* as the fundamental principles of our formalization. Concord simply means conformity with the norm, and discord, deviation from the norm.

Each genre has its own particular norm; for example, the style characteristic of scientific texts may seem disruptive if used in general newspaper writing. This view goes beyond the tenet of prescriptive rhetoric that the stylistic norm merely represents standard, 'good,' style, independent of the use of the text. Rather, we agree with Lanham (1974) that style should be viewed as the tailoring of a text to a specific audience and a specific situation; 'good' style can in fact be quite inappropriate.

The concept of norm will be useful as a means of anchoring the development of our stylistic rules. It is simply the usual, the typical, the stable and concordant. Indeed, discord, deviation from the norm, will play an essential role in our formalization of style, for it is our contention that style arises from the construction of patterns of

concord *and* discord, particular combinations of order and disorder that create an overall harmonious arrangement (Crystal and Davy 1969; Cluett 1976, 1990).

3.2.1 The Divisions of the Vocabulary. As the first step in constructing a vocabulary of style for computational use, we will look at the descriptive terms that stylists have developed over many centuries of use.

One of the most important rhetorical influences on sentence style that has been identified is syntactic *parallelism,* or *coordination:* the balancing of syntactically similar forms to achieve a harmonious stylistic effect, an idea dating back to the rhetorical *schemes of balance* (Corbett 1971, p. 463).

But if parallelism or coordination plays an important role in style, so too does *subordination:* variations in the hierarchical structure of a sentence that can produce correspondingly different stylistic effects. One part of a sentence can *dominate* another. As with parallelism, the importance to style of syntactic hierarchy is also a well-established principle in classical rhetoric, evident in the schemes of *parenthesis, apposition,* and *climax* (Corbett 1971, pp. 466, 468, 476). Thus, modulations in coordinate and subordinate structure are major factors in achieving stylistic expressiveness.

A third major contributor to syntactic style is *linear ordering:* varying the order of components within a sentence to produce quite marked stylistic differences. The rhetorical terms of *anastrophe* and *parenthesis* (Corbett 1971, p. 466) attest to the stylistic importance of the ordering of sentence components.

Thus, we can see that three factors influencing syntactic style are:

- **Balance:** Parallelism of structure.

- **Dominance:** Structural hierarchy.

- **Position:** Linear ordering of structure.

We will adapt these factors from theoretical stylistics as the divisions of our vocabulary of style. Now, we will classify commonly used stylistic terms into these three groups in order to identify and abstract the general properties of style characteristic of each group.

3.2.2 Abstract Properties of Style. In the *balance* group, we classify the terms *symmetric construction* (Hendricks 1976), *serial sentence* (Kane 1983), *parallel sentence* (Kane 1983), *balanced sentence* (Kane 1983), and *intersentence coordination* (Crystal and Davy 1969). All of these terms suggest a stylistic imitation, the balancing of syntactically similar clauses. We also find evidence of stylistic parallelism below the clause level in the use of terms such as *paired adjectives* (Kane 1983), *balanced phrases* (Crystal and Davy 1969), and *intrasentence coordination* (Hendricks 1976). As well as identifying similar structures, stylists have recognized characteristic *asymmetric constructions* (Hendricks 1976) or syntactic *counterpoint* (Hendricks 1976). These types of structures are most often *interrupted* or *convoluted* sentences (Kane 1983). For inclusion in our vocabulary, we will choose the most representative terms in the balance group: intersentence coordination and intrasentence counterpoint.

There are a variety of terms that can be classified into the *dominance* group. First, there are sentences that have one trivially dominant clause. These are the *simple sentences* (Kane 1983) that have only one main, or central, clause with no subordinate clauses. Next, there are many and varied types of sentences built up around a central, dominant clause. The *complex sentence* (Kane 1983) has a central clause and at least one dependent clause. The *loose sentence* (Kane 1983) has a central clause followed by

a subordinate clause. The *centered sentence* (Kane 1983) has a central clause flanked by subordinate clauses. The *cumulative sentence* (Kane 1983) has a central clause accompanied by a series of appositive, modifying, or absolute constructions. Finally, there are more-complicated sentences with more than one dominant clause. The *compound-complex sentence* (Kane 1983) can be much more elaborate than a simple compound sentence, as it contains at least two independent clauses and at least one dependent clause. For inclusion in our vocabulary, we abstract from these three types of hierarchical terms to identify trivial single-clause dominance, complex single-clause dominance, and multiple-clause dominance.

The *position* group appears to contain the largest number of terms, as any basic term can be qualified to an arbitrarily fine degree by the exact intrasentence position to which it refers. We will use only three qualifiers, associated with *initial, medial,* and *final* positions (Quirk et al. 1985). At the sentence level, we can have any word or phrase in initial, medial, or final position. Within the sentence, there are almost infinite variations of position terms. A small sample includes *postpositive adjectives,* and *pre-verbal, post-verbal,* or *post-clause adverbs* (Crystal and Davy 1969), and complex *premodification* and *postmodification* (Crystal and Davy 1969). We will generalize from these many subtle variations and define terms that have either a concordant or discordant effect, at a particular position in a sentence, according to whether or not their usage is normal.

From this grouping of terms into the three divisions of balance, dominance, and position, we have abstracted general properties of style that we want to describe in our vocabulary. The *abstract elements of style* that we will now define will classify the unstructured mass of stylistically significant sentence types into groups of sentences with similar stylistic properties.

3.3 Abstract Elements of Style

We will present the abstract elements in three groups, according to their properties of *balance, dominance,* and *position,* which we now define formally as follows:

Balance: A balance term characterizes a stylistic effect created by the juxtaposition of similar or dissimilar sentence components.
For example, *parallelism* in sentence structure is the juxtaposition of syntactically similar components.

Dominance: A dominance term describes a stylistic effect created by the particular hierarchical structure of a sentence.
In a *simple* sentence, for example, there is an effect of simplicity associated with the single and therefore, by default, dominant clause.

Position: A position term describes a stylistic effect created by the particular placement of a syntactic component within a sentence.
For example, in English, a *postmodifying* adjective, as in *the house ablaze,* can be more emphatic than the more usual *premodifying* type, as in *the blazing house.*

Now we shall propose a set of stylistic terms, correlated with this classification, that will make explicit those abstract stylistic properties that are now only implicit in existing terminology. These stylistic terms will be based on effects of *concord* and *discord,* which we formally define as follows:

Concord: A stylistic construction that conforms to the norm for a given genre.

Discord: A stylistic construction that deviates from the norm.

We will now use these notions to define abstract elements that are related, in turn, to effects of balance, dominance, and position.

3.3.1 Balance Elements. The first group of abstract elements is related to effects of balance within a sentence. Here, we will be looking at relationships that tend to either perturb or reinforce the balance of a sentence. The first and simplest type of balance element is a *homopoise* ("same weight"):

Homopoise: A sentence with interclausal coordination of syntactically similar components.

In other words, there are one or more stylistic 'shapes' in a homopoisal sentence, each contributing the same type of effect to the concordant parallelism of the sentence. In the example below, two very simple clauses, with identically shaped parse trees, are conjoined:

    13. The style was formed and the principles were acquired.[9]

Sentences that are more complex can have their balance interrupted or perturbed by a *heteropoisal* ("different weight") component:

Heteropoise: A sentence in which one or more parenthetical components are syntactically 'detached' and dissimilar from the other components at the same level in the parse tree.

It is difficult to formally characterize this class of structures and the underlying notion of a component being 'detached.' Quirk et al. (1985) also rely on this word to define what they call *disjuncts* but make no attempt to define it:

> Disjuncts ... have a superior role as compared with the sentence elements; they are syntactically more detached and in some respects 'superordinate,' in that they seem to have a scope that extends over the sentence as a whole. (Quirk et al. 1985, p. 613)[10]

Similarly, Nunberg (1990), in defining the related idea of the *delimiter comma*, can do no better than to characterize the class "in a rough way":

> There is the class of elements *delimited* by commas, either at both ends (when the elements occur clause-internally) or at one end (roughly, when the elements are either clause-initial or -final) [p. 36]. ... It is obviously not possible here to provide an analysis of the syntax and semantics of each of these constructions, much less to try to characterize them in such a way as to say what they have in common. In a rough way, however, we observe that they all involve constituency

---

9 Adapted from The *Manchester Guardian Weekly*, 14 February 1988, p. 15.
10 An example of a disjunct is the interrupting prepositional phrase *in all frankness* in the sentence *Your son is not, in all frankness, succeeding in his present job* (Quirk et al. 1985, p. 612).

problems, so that they do not behave as constituents, for example, under such tests as verb-phrase ellipsis. What is more, all of them can be characterized as supplying material that is communicatively supererogatory to the bare propositional content of the lexical clause in which they appear. ... This is all very rough, but it does suggest that the content of these comma-delimited elements plays a distinctive role in the representation of the [rhetorical] argument structure of the text. [pp. 38–39]

Despite the difficulty of formal characterization, we agree with Quirk et al. and Nunberg that there is an intuitively clear class of constructions here, and, like these authors, hope that it can be conveyed to the reader by sufficient use of examples.

We divide the class of heteropoise sentences into three types, depending upon the position of the parenthetical component: a heteropoise may be *medial, initial,* or *final.* Moreover, we will distinguish between two subtypes at each position: *concordant* and *discordant.*

Medial heteropoise: A heteropoise in which the parenthetical component is in medial position.

> **Concordant medial heteropoise:** A medial heteropoise in which the parenthetical component is cohesively linked to the rest of the sentence. (This notion will be made precise when we have introduced the *primitive elements* in Section 3.4.)
>
> **Discordant medial heteropoise:** A medial heteropoise in which the parenthetical component is *not* cohesively linked to the rest of the sentence.

In example 14 below, the relative clause *which brought no protests* interrupts the main clause, yet, as a postmodifier, it is still a part of the subject noun phrase and so the sentence is a *concordant medial heteropoise.* But now consider the medial adverbial phrase *according to a company spokesman*; as a sentence modifier, it is quite detached from the rest of the sentence and so the sentence is simultaneously a *discordant medial heteropoise*:

14. The measure, <u>which brought no protests</u>, was decided, <u>according to a company spokesman</u>, because of the dangers to which these employees are exposed in travelling to particularly exposed subtropical countries.[11]

Initial heteropoise: A heteropoise, concordant or discordant, in which the parenthetical component is in initial position.

Sentence 15 contains an initial parenthesis, an appositive noun phrase, that is cohesively linked to the rest of the sentence because it refers to the subject of the main clause. Thus, the sentence is a *concordant initial heteropoise*:

15. <u>The heir to a fortune</u>, her friend did not care about passing examinations.[12]

---

11 *Manchester Guardian Weekly*, 21 February 1988, p. 14.
12 Quirk et al. (1985, p. 1314).

Final heteropoise: A heteropoise, concordant or discordant, in which the paren-
thetical component is in final position.

Sentence 16 contains a final parenthesis, an 'extraneous' phrase, that is distinct from
the main clause, as it is not cohesively linked to the rest of the sentence. Thus, the
sentence is a *discordant final heteropoise*:

16. Stephen and Jennifer are not going to buy the house, <u>according to a
    spokesman</u>.

We saw in 14 that a heteropoise can contain more than one parenthetical, and one
of them might be concordant while another is discordant. Likewise, it can be the case
that one parenthetical is, say, medial, while another is initial or final. The following
example shows two successive medial parentheticals rather stridently interrupting the
main text—the effect is heightened by the use of dashes—as well as a final parenthetical
in the same sentence:

17. The idea of combined French and British patrols by nuclear
    submarines—<u>a proposal once made by the Social Democrat
    leader David Owen</u>—<u>let alone the plan for "sharing the work"
    where the targets and missiles carried by these submarines are
    concerned</u> was not even raised, <u>according to a French military
    spokesman</u>.[13]

3.3.2 Dominance Elements. The second type of abstract element deals with stylis-
tic dominance, which is concerned with the hierarchical structure of a sentence. A
common type of dominance element is the *monoschematic*, a very simple sentence:

Monoschematic: A sentence with a single main clause with simple phrasal sub-
ordination and no accompanying subordinate or coordinate clauses.

Here is an example of a canonical monoschematic sentence:

18. Posterity has not been kind to him.[14]

The most common dominance element in the texts that we have studied is the
centroschematic:

Centroschematic: A sentence with a central, dominant clause with one or more
of the following optional features: complex phrasal subordination, initial
dependent clauses, final dependent clauses.

Centroschematic sentences can be quite varied in structure, but their shared charac-
teristic is a predominant component that serves as the organizational center for all
other components. Such structures are built with subordination and coordination. In

---

13 *Manchester Guardian Weekly*, 7 February 1988, p. 13.
14 Adapted from The *Manchester Guardian Weekly*, 14 February 1988, p. 15.

the first example below, the main clause is supported by a complex, but subordinate and coordinate, relative clause structure:

19. Neither these devices nor the cramped viewing rooms <u>which are too narrow and whose ceilings are much too low for the big altarpieces</u> manage to spoil the works.[15]

Two subordinate clauses, the first of them adverbial and the second of them relative, accompany but do not dominate the main clause in the following sentence:

20. Not all that long ago, the famous collector Charles de Bestegui, <u>when unable to get hold of certain paintings</u>, was quite prepared to adorn the walls of his Venetian palace with copies, <u>which happily rubbed shoulders with his numerous genuine canvases</u>.[16]

Finally, the complex but imitative postmodification in the next example incorporates a substantial amount of information without weakening the dominance of the main clause:

21. Silvia, <u>a commanding woman in her 50's</u>, <u>a shrew falsely mellowed by religion</u>, promptly organised prayer sessions on the lines of Tupperware meetings.[17]

Given that we have a progression in complexity from monoschematic to centroschematic sentences, a natural extension is to the *polyschematic*:

Polyschematic: A sentence with more than one central dominant clause and at least one dependent clause.

Such sentences occur much less frequently than the monoschematic or centroschematic varieties, at least in the corpus used in this research. However, a sentence with obviously disparate components occasionally occurs, as in the following example. There are two dominant clauses (*we could think* ... and *we should not forget* ...) and a dependent clause (*if we consider the progress* ...):

22. If we consider the progress already achieved, the opposition that had to be overcome, for example, in order to open schools for girls, and the fact that Saudi Arabia is less than 60 years old, we could think that time will permit resolving the contradictions between the most liberal aspirations of one part of society and the ulemas' determination to keep the country as it is, and we should not forget to mention the Islamic fundamentalist movements which are threatening Saudi Arabia.[18]

---

15 *Manchester Guardian Weekly*, 14 February 1988, p. 15.
16 *Manchester Guardian Weekly*, 7 February 1988, p. 14.
17 Adapted from The *Manchester Guardian Weekly*, 7 February 1988, p. 16.
18 Adapted from The *Manchester Guardian Weekly*, 14 February 1988, p. 14.

3.3.3 Position Elements. The third group of abstract elements that we will look at are the position elements. The most common types of position element describe *concordant* or *discordant* stylistic effects in particular positions. Thus, we have *initial concord, medial concord*, and *final concord*. Similarly, we have a range of *discord* elements. The definitions of these elements are closely tied to the primitive-level descriptions of our vocabulary, so we will delay their precise characterizations until Section 3.5, after the primitive-level descriptions are introduced. Here, it will be sufficient for the reader to know only that the concords describe constructions, at a particular position in the sentence, that conform to normal usage while the discords describe constructions that deviate from the norm.

In addition to effects created by a single concord or discord in a particular position within the sentence, we can observe other kinds of effects pertaining to the relationships between concords and discords:

Resolution: A shift in stylistic effect that occurs at the end of a sentence and is a move from a relative discord to a stylistic concord.

We can see an example of a resolution in 23 below. There is an initial discord, created by the unusual placement of the adverb *entirely*, which is not normally found in the initial position and is not cohesively linked to the rest of the sentence (cf. Section 3.4.3 below). However, there is a final concord, as the subsequent main clause contains no such incongruities:

23. Entirely in the spirit of protective support, could we suggest you pass on an appropriate comment to the personnel concerned.[19]

And the complementary effect is dissolution:

Dissolution: A shift in stylistic effect that occurs at the end of a sentence and is a move from a relative concord to a stylistic discord.

We have already seen an example of a dissolution in one of the Biblical texts of Section 1.1, repeated below as 24. A strong initial concord, created by imitative clauses, is set against a final discord, produced by an uncommon syntactic inversion:

24. And the rain descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it.[20]

3.4 Primitive Elements of Style

3.4.1 Two Kinds of Stylistic Analysis. Now that a set of abstract stylistic elements has been defined, we must next consider how to use them in practical stylistic analysis. So far, these elements are too general to apply to the interpretation of an arbitrary sentence. We must relate these *stylistic* elements to more-basic *syntactic* elements, whose stylistic characteristics are more specific and concrete. In our reading of the literature

---

19 Quirk et al. (1985, p. 652).
20 Matthew 7:27, Revised Standard Version (1952).

of stylistics, we observed that two analyses of stylistic structure are possible: *connective* and *hierarchic* orderings:

Connective ordering: The result of cohesive bonds drawing together components in a linear ordering.

Hierarchic ordering: The result of bonds of subordination and superordination drawing together components in a nested ordering.

These two complementary kinds of analysis are implicit in the work of most stylists and rhetoricians, such as Cluett (1976) and Bureau (1976) (Neil Randall, personal communication).

We will now use these notions of connective and hierarchic orderings to guide the definition of more-primitive stylistic elements that provide a precise syntactic basis to the vocabulary, yet also allow a mapping to the abstract elements.

3.4.2 The Role of Corpus Studies in the Classification of the Primitive Elements of Style. We believe that to establish stylistic classifications and construct a practical stylistic grammar, one must undertake corpus studies in the style of Cluett (1976) and Biber (1988, 1989) to support the exact classifications of stylistic effects. But we also agree with Crystal and Davy that:

> It is perhaps worth emphasising right away that the first step in any stylistic analysis must be an intuitive one. The stylistician is on precisely the same footing as anyone else here: he notices a linguistic feature which he feels to be stylistically significant. The difference between his approach and that of the untrained observer is that he will have a clearer idea of what is likely to be significant, and will know what to do with his observations once they are made. This last point is the stylistician's main competence: he is able to interrelate his observations within the framework of some theory, and thus piece together any general pattern of linguistic variation which may exist. (Crystal and Davy 1969, p. 12)

Thus, although we have built our grammar of style upon accepted theory, it is currently a prototype for the formal representation of stylistic knowledge, not yet developed as a tool for practical stylistic analysis.[21] For, like Crystal and Davy, our initial objective is to provide a methodology, a vocabulary and apparatus, that will allow others to integrate our approach with analytic procedures to construct useful formalizations of style:

> All we can do is systematically point to certain significant facts in the language being analysed, suggest some theoretical principles which will account for the occurrence of these facts, emphasise the need for further analysis to validate or refute these preliminary soundings, and illustrate a procedure which will allow people to do this. (Crystal and Davy 1969, p. 13)

---

21 See Section 6 for a discussion of work in progress.

In the manner of Crystal and Davy, we agree that statistical verification must follow such research, but we believe that, initially, we can use well-motivated intuitions about stylistic frequencies to classify our stylistic elements:

> Such a [statistical] method would be prerequisite for any serious re-search work, but it is too detailed for our present purposes. Conse-quently we make use of the alternative method of expression, making our statements of frequency in more informal terms, using such quan-tifiers as 'rarely', 'commonly', 'often', 'very often', and so on. This range of adverbials in English cuts up a continuum of frequency very clearly, with very little overlap, and is readily intelligible. It should not be forgotten, however, that any statements of relative frequency in these terms can be referred if necessary to the precise statistical situation which underlies them. (Crystal and Davy 1969, p. 22)

3.4.3 Primitive Stylistic Elements.

Primitive stylistic elements are individual con-stituents that have a particular stylistic effect. But what should we use as the basis for the definition of 'stylistic effect'? We have said that the two analyses of sentence structure, connective and hierarchic, can provide the appropriate guidance. We will define the stylistic effect of an individual component in terms of its contribution to each ordering:

- For the connective view, a component acquires its stylistic effect from its degree of *cohesiveness*, its *bonding* with other components in the sentence.

- For the hierarchic view, a component acquires its stylistic effect from its degree of *subordination*, its *dependence* on other components in the sentence.

We introduce the terms *conjunct, antijunct, subjunct,* and *superjunct* (to be defined in the following sections) and we use superscripts on all the terms to indicate the degree of connectivity or subordination. In the connective view, we classify syntactic com-ponents as either *conjunct^4* (excessively connective), *conjunct^3* (strongly connective), *conjunct^2* (moderately connective), *conjunct^1* (mildly connective), and *conjunct^0* (neu-tral). Similarly, we use the terms *antijunct^0* through *antijunct^4* to indicate increasingly disconnective effects; *conjunct^0* and *antijunct^0* are the same.

There is a complementary vocabulary of primitive elements for the hierarchic view. The stylistic effects of syntactic components are correlated with the degree of subordi-nation or superordination; the classifications are analogous to the connective: *subjunct^4* through *subjunct^0* (decreasingly subordinate) and *superjunct^0* through *superjunct^4* (in-creasingly superordinate); *subjunct^0* and *superjunct^0* are the same.

Connective primitive elements. In assigning connective effects to syntactic compo-nents, we adapt Halliday and Hasan's (1976) work on cohesion. Although Halliday and Hasan focused on definitions of cohesion in whole texts, we can apply these definitions to cohesion *within* the sentence, for, as they point out:

> Since cohesive relations are not concerned with structure, they may be found just as well within a sentence as between sentences. They attract less notice within a sentence, because of the cohesive strength of grammatical structure; since the sentence hangs together already,

the cohesion is not needed in order to make it hang together. But the
cohesive relations are there all the same. (Halliday and Hasan 1976,
p. 8)

Halliday and Hasan enumerate five types of cohesion: *substitution, ellipsis, reference,
conjunction*, and *lexical cohesion*. We will use all of these except lexical cohesion and
will add a new factor, *interpolation*, that works against cohesion.

Halliday and Hasan define *substitution* as the "replacement" (p. 88) of one item
by another. We suggest that such intrasentence phenomena as apposition and extra-
position can be considered as types of substitution. For example, in 25, either of the
two postmodifying noun phrases could substitute for the head noun *Silvia*:

25. Silvia, a commanding woman in her 50s, a shrew falsely mellowed by
    religion, promptly organized prayer sessions on the lines of Tupperware
    meetings.[22]

And, in 26, the clausal subject *for anyone to escape* could substitute for the anticipatory
pronoun *it*:

26. It was considered impossible for anyone to escape.[23]

Ellipsis is defined as the omission of an item. Many forms of ellipsis are possible
within the sentence, including the omission of elements of clause structure. In the
following sentence, the ellipsis is marked by "___":

27. I'll gladly pay for the hotel, if you will ___ for the food.[24]

Items that have the cohesive property of *reference* are not interpreted semantically
in their own right, but make reference to something else for their interpretation (Halli-
day and Hasan 1976, p. 31). In English, such items are personal pronouns, demonstra-
tives, and comparatives. These 'directive' elements presuppose the existence of some
other element from which information is to be retrieved; whether or not the other
element is within the same sentence, the referential item will be cohesive.

Conjunctive elements are defined as:

> elements [that] are not cohesive in themselves but indirectly, by virtue
> of their specific meanings; they are not primarily devices for reaching
> out into the preceding (or following) text, but they express certain
> meanings which presuppose the presence of other components in the
> discourse. (Halliday and Hasan 1976, p. 226)

Halliday and Hasan cite the conjunction *and*, certain adverbs, and certain prepositional
expressions as instances of conjunction. Many of these elements can promote cohesion
within the sentence.

To the principles of cohesion described above, we will add *interpolation*, a discon-
nective relation that works *against* cohesion. Interpolated elements are certain instances

---

22 Adapted from The *Manchester Guardian Weekly*, 7 February 1988. Translated from *Le Monde*.
23 Quirk et al. 1985, p. 1392.
24 Quirk et al. 1985, p. 907.

of parenthetical constructions, those that display none of the forms of cohesion listed earlier.[25] In the examples below (repeated from Section 3.3.1), we see several different types of interpolation:

28. The measure, <u>which brought no protests</u>, was decided, <u>according to a company spokesman</u>, because of the dangers to which these employees are exposed in travelling to particularly exposed subtropical countries.[26]

29. <u>The heir to a fortune</u>, her friend did not care about passing examinations.[27]

30. Stephen and Jennifer are not going to buy the house, <u>according to a spokesman</u>.

The cohesive relations described above can be ranked from the most to the least cohesive. Halliday and Hasan (pp. 226–227) consider substitution, including ellipsis, to be the most strictly cohesive relation, followed by reference, and then conjunction. We adopt this ranking, and so we classify intrasentential substitution and ellipsis as strongly connective (*conjunct³*), reference as moderately connective (*conjunct²*), and conjunction as mildly connective (*conjunct¹*). We classify interpolation as disconnective (*antijunct²*).[28]

As an illustration of our method of classification of connective primitive elements, we will use the following sentences, in which the underlined components exhibit distinct stylistic effects. In the first sentence, the postmodifying construction is a strongly connective, *conjunct³* reduced relative clause, as it is an instance of ellipsis for which the presupposed items are *who is*. The postmodifying relative clause in 32 is a moderately connective, *conjunct²* referential element, as the relative pronoun 'points' to the preceding noun phrase *the man*. Lastly, the postposed adjectival in 33 is a disconnective, *antijunct* instance of interpolation, as it is detached from the rest of the sentence and has none of the cohesive properties. (The property of ellipsis is not present, as a postposed adjective, unlike a reduced relative clause, does not have the same element of presupposition of specific missing items.):

31. The man ___ <u>walking into the room</u> is unfit for this task.

32. The man <u>who is walking into the room</u> is unfit for this task.

33. A man <u>always timid</u> is unfit for this task.[29]

Hierarchic primitive elements. In assigning a hierarchic classification to a syntactic component, we adapt Halliday's (1985) work on *subordination* in functional grammar and Quirk et al.'s (1985) use of the term *superordination*.

---

25 Another type of structure that seems to work against cohesion is a disruption in normal linear ordering, as the syntactic inversion at the end of the now-familiar text, *And the rains descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it.* However, we have not yet incorporated this feature into our formalization of style.
26 *Manchester Guardian Weekly*, 21 February 1988, p. 14.
27 Quirk et al. (1985, p. 1314).
28 We do not use here the full range of connectivity that we defined above; however, some extensions of the theory by Green (1992a, 1992b) do.
29 Quirk et al. (1985, p. 1295).

Halliday distinguishes two types of subordination: *embedding* (or *rank-shifting*) and *hypotaxis*. These phenomena will be used to classify the hierarchic primitive elements in our stylistic grammar. *Embedding* is described as:

> the 'rank shift' by which a clause or phrase comes to function within the structure of a group, like *who came to dinner* in *the man who came to dinner*. (Halliday 1985, p. 219)

There are various types of embeddings, including prepositional phrases, as in *John might arrive <u>as early as tomorrow</u>,* but Halliday focuses on those within the nominal group:

> [The embedded prepositional phrase] is the only instance of embedding other than in a nominal group. All other embedding in English is a form of nominalization, where a group, phrase, or clause comes to function as part of, or in place of (i.e., as the whole of), a nominal group. (Halliday 1985, p. 187)

Examples of embedded subordination are finite and nonfinite clauses[30] that function as postmodifiers, as in *the house <u>that Jack built,</u>* or as the head of a nominal group, as in *<u>For Jack to build a house</u> would be the best thing* (Halliday 1985, p. 220).

Halliday defines *hypotaxis* as:

> the binding of elements of unequal status. The dominant element is free, but the dependent element is not. [p. 198] In a hypotactic structure the elements are ordered in dependence, and this ordering is largely independent of the sequence. [p. 199]

There are many types of hypotaxis, at the clause, phrase, and word levels. Examples of hypotactic subordination include a dependent clause following a dominant clause, preceding the dominant, enclosed in the dominant, or enclosing the dominant:

34. You never can tell <u>till you try</u>.[31]

35. <u>If wishes were horses</u>, beggars would ride.[32]

36. Picture, <u>if you can</u>, a winkle.[33]

37. <u>He might</u>, he said, <u>finish it himself</u>.[34]

It should be emphasized that the two types of subordination are quite different:

> It is important to distinguish between embedding on the one hand and ... hypotaxis on the other. Whereas hypotaxis [is a relation] BETWEEN clauses (or other ranking elements), embedding is not. Embedding is

---

30 If the clause is relative, then it must be defining (restrictive) to be an embedded clause; a non-defining relative clause is an instance of hypotaxis (to be defined below).
31 Halliday 1985, p. 200.
32 Halliday 1985, p. 200.
33 Halliday 1985, p. 200.
34 Halliday 1985, p. 200.

> a mechanism whereby a clause or phrase comes to function as a constituent WITHIN the structure of a group, which itself is a constituent of a clause. (Halliday 1985, p. 219)

Consequently, the stylistic effects of embedding and hypotaxis will be subtly different. Because an embedded structure is a *constituent* of the dominant structure, we will classify embeddings as strongly subordinate, *subjunct*[3]. On the other hand, in hypotaxis, one component is dependent on another, but "in no sense is it a constituent part of it" (Halliday 1985, p. 219), so we will classify hypotactic structures as only mildly subordinate, *subjunct*[1].

To the factors of embedding and hypotaxis, we add another factor, *superordination*, that accounts for structures that are somehow 'superior' to the main clause.[35] That such phenomena exist is supported by Quirk et al.'s (1985) definition of *disjunct*, which we repeat here from Section 3.3.1:

> Disjuncts ... have a superior role as compared with the sentence elements; they are syntactically more detached and in some respects 'superordinate,' in that they seem to have a scope that extends over the sentence as a whole. (Quirk et al. 1985, p. 613)

Disjuncts can be divided into two main classes, *style* disjuncts and *content* disjuncts:

> Style disjuncts convey the speaker's comment on the style and form of what he is saying, defining in some way under what conditions he is speaking as the 'authority' for the utterance. Content disjuncts (also known as attitudinal disjuncts) make observations on the actual content of the utterance and its truth conditions. (Quirk et al. 1985, p. 615)

Examples of these superordinate elements are underlined in the following sentences (the first three contain style disjuncts, the fourth, a content disjunct):

38. <u>Frankly</u>, I am tired.[36]

39. <u>If I may say so without offence</u>, your writing is immature.[37]

40. There were twelve people present, <u>to be precise</u>.[38]

41. <u>Even more important</u>, she has control over the finances of the party.[39]

In our grammar, we classify these types of superordinate elements as *superjunct*.

As an illustration of our method of classification of hierarchic primitive elements, we will use the following sentences, which display distinct stylistic effects due to the different primitive classifications of the underlined components. In the first sentence below, the postmodifying restrictive relative clause is a strongly subordinate, *subjunct*[3] instance of embedding. The medial adverbial clause in 43 is a mildly subordinate,

---

35 We do not consider the case of a main clause being superordinate to a dependent clause (Quirk et al. 1985, pp. 988–990). We are concerned only with structures that are superordinate to the matrix clause.
36 Quirk et al. 1985, p. 615.
37 Quirk et al., p. 615.
38 Quirk et al., p. 616.
39 Quirk et al., p. 622.

subjunct[1] instance of hypotaxis, as it is a causal-conditional clause (Halliday 1985, p. 214). Lastly, the initial adjective in 44 is a superordinate, *superjunct* instance of a style disjunct.

42. The money <u>which was our reward</u> has been useful.

43. The money, <u>as it was our reward</u>, should not be used for frivolous purposes.

44. <u>True</u>, the money has been useful.

3.5 Concords and Discords

In Section 3.3.3, we gave an informal description of the abstract elements of *concord* and *discord*. As these elements are closely tied to the primitive-level descriptions of our vocabulary, we delayed their precise characterizations until after the primitive elements of style were introduced. Now, we can define the nature of concords and discords for both the connective and hierarchic views. In both views, the concords represent constructions, at a particular position in the sentence, that conform to normal usage, while the discords describe constructions that deviate from the norm.

In the connective view, the concords are associated with constructions that are connective, for we assume that cohesion is the default usage, while the discords are disconnective structures, as a lack of cohesion is less usual. Thus, a concord, which may be *initial*, *medial*, or *final*, is produced by a *conjunct* construction; a discord, which may also appear in any of the three positions, is caused by an *antijunct* construction.

For the hierarchic view, concords are associated with structures that display a 'normal' degree of subordination or superordination, while discords are produced by excessive subordination or excessive superordination. Excessive subordination and superordination are difficult to characterize, but, as a first attempt, we propose that nesting of dependent structures to a depth greater than three will be considered discordant subordination and that 'detachment' of superordinate structures by 'intrusive' punctuation (e.g., long dashes) will be considered discordant superordination.

4. A Stylistic Grammar

4.1 A Methodology for Developing a Stylistic Grammar

So far, we have constructed a vocabulary of style. We can now define a method for using the vocabulary to systematically build up a syntactic stylistic grammar for both the connective and hierarchic views of style.

To construct the stylistic grammar, we will use a methodology whose steps are as follows:

- the classification of primitive stylistic elements;

- the correlation of the stylistic effects of these elements with the abstract stylistic elements; and

- the correlation of patterns of these abstract elements with specific stylistic goals.

The grammar to be constructed in this manner will be stratified with internal branching, as illustrated in Figure 1. As the figure shows, at the bottom level there are three branches—lexical, syntactic, and semantic—each with its own vocabulary
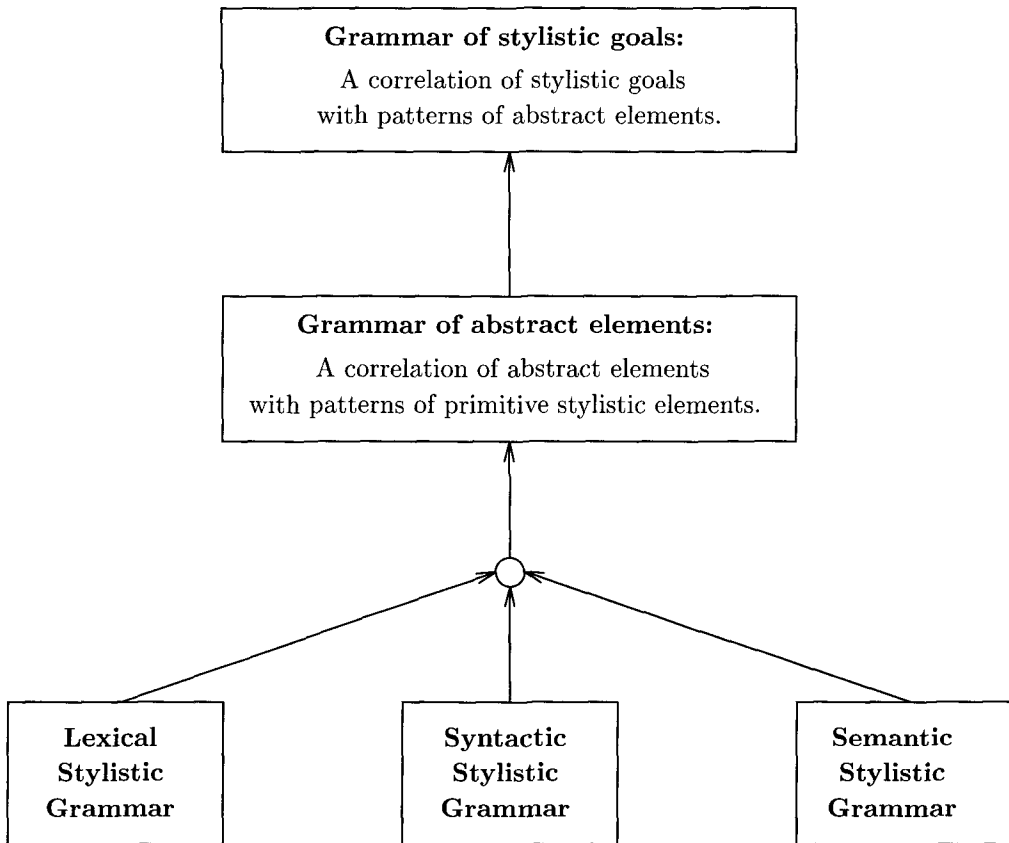
Figure 1
A stylistic grammar.

of primitive stylistic elements and rules for combining them. At the central level, we use a single vocabulary of abstract elements, stylistic terms that are maximally expressive. Rules relate these abstract elements to patterns of lexical, syntactic, and semantic primitive elements. This level is the unifying core of the grammar; as we will note later, the same abstract elements can describe both English and French style, and both syntactic and semantic aspects of style. Finally, at the top level, rules correlate individual stylistic goals with patterns of abstract elements. Together, these levels form a language-independent[40] schema for a goal-directed grammar.

4.2 Grammar of Primitive Elements

The stylistic grammar that we have outlined in the previous section shows how lexical, syntactic, and semantic aspects of style would be integrated within one grammar. Some work has already been done on the development of the lexical and semantic branches (see Sections 6.1 and 6.2), but here we will focus on syntactic style. We will now trace the development of the syntactic component of the stylistic grammar, beginning with the most detailed and most closely syntactic level, the grammar of primitive elements. As we proposed earlier, we will consider two analyses of sentence structure, connective and hierarchic, each with its own stylistic classifications. In this base-level grammar of

---

40 We expect that a particular stylistic grammar will be applicable to a family of stylistically similar
   languages.

primitive elements, the rules are based on syntactic and stylistic rules of Crystal and Davy (1969), Kane (1983), and Quirk et al. (1985).

In the full grammars given by DiMarco (1990) and subsequently by Hoyt (1993), each syntactic category is given both a connective and a hierarchic primitive classification by corresponding rules. Here, we will present only selected rules, mainly those associated with the stylistic effects of various types of premodification and postmodification in the nominal group. We will begin with the elementary components and work up to full sentences. In the presentation of the grammar rules, we will first give the connective rules, then their hierarchic counterparts. Often, we will annotate rules with relevant examples. (Explanatory notes on the terminology are given in Appendix A.)

Adjectivals. Adjectivals and adverbials are the simplest classes of primitive elements. To demonstrate that even the smallest syntactic components have individual stylistic effects, we give selections from the rules for the adjectivals.

Connective view. Adjectivals can be classified within a range of connectivity. In the moderately connective case, there are a number of alternatives, two of which are shown below. The first alternative, the genitive form, is a referential element, as it presupposes the existence of the noun to which it refers. Whether or not this noun is in the same sentence, a referential personal pronoun is cohesive and so it is classified as a *conjunct*[2] element. The second alternative, a demonstrative determiner, is also classified as moderately connective because it too is a referential element.

Rule
conjunct[2] adjectival $\longrightarrow$

   premodifying genitive
      his religious works

   demonstrative determiner
      this substantial selection

The less-conjunct case has only one alternative, an adjective. An adjective is usually used as a premodifier of a noun, so that it presupposes the presence of the noun. However, it is not a directive, referential element, but rather a conjunctive one. Thus, it is classified as *conjunct*[1].

Rule
conjunct[1] adjectival $\longrightarrow$

   adjective
      medieval iconographic subject

Hierarchic view. Adjectivals are generally hypotactic rather than embedded elements. That is, they are linked to an element (the noun) of different status that could stand as an independent whole, whereas the adjectival could not. Therefore, we will classify most adjectivals as *subjunct*[1], mildly subordinating. However, nominal adjectives, which have been rank-shifted, are instances of embedding and so are *subjunct*[3], more strongly subordinating.

Rule
  subjunct[1] adjectival $\longrightarrow$

    premodifying genitive
      his religious works

    demonstrative determiner
      this substantial selection

    adjective
      medieval iconographic subject

Rule
  subjunct[3] adjectival $\longrightarrow$

    noun
      fake cloth architectural backdrops

Premodification. Crystal and Davy's (1969) 'stylistic' grammar, although elementary, rightly recognized the stylistic importance of premodification and postmodification of the nominal group. Our syntactic stylistic grammar is built upon the stylistic effects created by these types of modification. We first describe selected rules of *premodification*.

Connective view. The rules for moderately connective, *conjunct[2]*, and less connective, *conjunct[1]*, *premodification* are self-explanatory.

Rule
  conjunct[2] premodification $\longrightarrow$

    conjunct[2] adjectival

Rule
  conjunct[1] premodification $\longrightarrow$

    conjunct[1] adjectival

The following rule for non-connective, *conjunct[0] premodification* has three alternatives, an adverbial, a participle, and no premodification. Premodifying adverbials and participles do not display any of the explicit forms of cohesion but neither do they work against cohesion. They are not connective but neither are they disconnective.

Rule
  conjunct[0] premodification $\longrightarrow$

    adverbial
      increasingly arresting self-portraits

    participle
      arresting self-portraits

    no premodification

The following rule for disconnective, *antijunct²* *premodification* has only one alternative, a reduced sentence. A reduced sentence can be considered to be an instance of interpolation, so that it works against cohesion.

Rule
antijunct² premodification $\longrightarrow$

    reduced sentence
        I visited his <u>pop-down-for-the-weekend</u> cottage.

We can now begin to introduce primitive stylistic categories that will be used to build up the definitions of the corresponding abstract elements that we saw in earlier sections. When premodification is conjunct, it is classified as *centroschematic premodification*. In turn, this type of premodification is *concordant*, for it is associated with cohesive effects, which we take to be normal usage.

Rule
centroschematic premodification $\longrightarrow$

    conjunct³ premodification

    conjunct² premodification

    conjunct¹ premodification

    conjunct⁰ premodification

Rule
concordant premodification $\longrightarrow$

    conjunct³ premodification

    conjunct² premodification

    conjunct¹ premodification

    conjunct⁰ premodification

When premodification is antijunct, then it is *discordant*, for there is a degree of disconnectivity that works against cohesion; it is therefore not normal usage.[41]

Rule
discordant premodification $\longrightarrow$

    antijunct premodification

---

41 For reasons of brevity, we will occasionally use a single term, such as *antijunct*, to stand for all degrees of an element.

Hierarchic view. In the full grammar, we define detailed rules for the various subjunct and superjunct types of premodification. Here, we will show only selected rules.

The following rule for strongly subordinating, *subjunct³ premodification* has four alternatives. The first is self-explanatory. The other three, reduced sentences, adverbials, and participles, are embedded, rank-shifted constructions, which are therefore classified as *subjunct³*.

Rule

subjunct^3 premodification $\longrightarrow$

     subjunct^3 adjectival

     reduced sentence

     adverbial

     participle

The following rule for less subordinating, *subjunct¹ premodification* is self-explanatory.

Rule

subjunct^1 premodification $\longrightarrow$

     subjunct^1 adjectival

We include the rule below to account for the case of no premodification, which is neither subordinating nor superordinating.

Rule

subjunct^0 premodification $\longrightarrow$

     no premodification

Centroschematic premodification is produced by subjunct or 'neutral' (*subjunct⁰*) premodifying constructions.

Rule

centroschematic premodification $\longrightarrow$

     subjunct^3 premodification

     subjunct^2 premodification

     subjunct^1 premodification

     subjunct^0 premodification

In the hierarchic view, discord is produced by excessive subordination. We will classify this type of discordant premodification as *subjunct⁴ premodification*. Now we can define the following rules for concordant and discordant premodification:

479

Rule
  concordant premodification $\longrightarrow$

   subjunct^3 premodification

   subjunct^2 premodification

   subjunct^1 premodification

   subjunct^0 premodification

Rule
  discordant premodification $\longrightarrow$

   subjunct^4 premodification

Postmodification. Continuing to build upon the stylistic features that Crystal and Davy (1969) judged important, we define complementary connective and hierarchic rules for *postmodification* in the full grammar and give selected rules here.

Connective view. The rule for *conjunct^3 postmodification* has three alternatives, a nominal group, a nonfinite clause, and a verbless clause.[42] As these constructions are instances of either substitution or ellipsis, they are strongly connective.

Rule
  conjunct^3 postmodification $\longrightarrow$

   nominal group
     Paul Jones, the distinguished art critic, died in his sleep last night.

   non-finite clause
     You will look in vain for any concrete measures emerging from this summit.

   verbless clause
     Norman Jones, then a student, wrote several best-sellers.

As a relative clause is a referential structure, it is classified as a moderately connective, *conjunct^2* shape.

Rule
  conjunct^2 postmodification $\longrightarrow$

   relative clause
     I do not trust a laboratory that will not insist on a qualified pharmacist
     being present at the point of distribution.

A postmodifying prepositional phrase is classified as mildly connective, *conjunct^1*,

---

42 We follow Quirk et al. (1985, p. 1314) in using this apparent contradiction in terms.

because it is a conjunctive element that presupposes the presence of the noun it modifies.

Rule
conjunct[1] postmodification $\longrightarrow$

>   prepositional phrase
>   *his long black cloak _with its purple beading and ornamentations of gold and precious stones_*

A postposed adjectival is an instance of interpolation, and therefore *antijunct,* as it is a detached construction that lacks any of the forms of cohesion.

Rule
antijunct[2] postmodification $\longrightarrow$

>   adjectival
>   *A man _always timid_ is unfit for this task.*

If postmodification is conjunct, then it is *centroschematic postmodification.*

Rule
centroschematic postmodification $\longrightarrow$

>   conjunct[3] postmodification
>
>   conjunct[2] postmodification
>
>   conjunct[1] postmodification
>
>   conjunct[0] postmodification

If postmodification is either conjunct and parenthetical or antijunct and parenthetical, then it is *heteropoisal.*[43,44]

Rule
heteropoisal postmodification $\longrightarrow$

>   conjunct postmodification *with* parenthesis
>
>   antijunct postmodification *with* parenthesis

Now we can define the various types of *concordant* and *discordant postmodification,* according to whether the postmodification is cohesive (and therefore concordant) or lacking cohesion (discordant).

---

43 We do not define all forms of parenthesis, but these can be easily enumerated by using, for example, the classifications given by Quirk et al. (1985).
44 The meaning of *with* in our grammar is explained in the notes on terminology in Appendix A.

Rule
   concordant postmodification $\longrightarrow$

   conjunct^3 postmodification

   conjunct^2 postmodification

   conjunct^1 postmodification

   conjunct^0 postmodification

Rule
   discordant postmodification $\longrightarrow$

   antijunct postmodification

Hierarchic view. In the rule for *subjunct^3 postmodification*, the alternatives are a restrictive relative clause, a restrictive nonfinite clause, a nominal group, and a prepositional phrase. All of these are instances of embedding, as they function as a part of the nominal group, so they are strongly subordinating elements.

Rule
   subjunct^3 postmodification $\longrightarrow$

      restrictive relative clause

      restrictive non-finite clause

      nominal group

      prepositional phrase

   In the rule for *subjunct^1 postmodification*, the alternatives are an adjectival, a nonrestrictive relative clause, and a nonrestrictive nonfinite clause. All of these are instances of hypotaxis, as none of them is a constituent of the nominal group, so they are classified as mildly subordinating elements.

Rule
   subjunct^1 postmodification $\longrightarrow$

      adjectival

      non-restrictive relative clause

      non-restrictive non-finite clause

Monoschematic postmodification is produced by simple subordinate forms, excluding clauses.[45]

Rule
monoschematic postmodification $\longrightarrow$

subjunct[3] postmodification *and* (noun phrase *or* prepositional phrase)

Centroschematic postmodification is produced by all subordinate forms, including clauses.

Rule
centroschematic postmodification $\longrightarrow$

subjunct[3] postmodification

subjunct[1] postmodification

subjunct[0] postmodification

In the hierarchic view, discord is produced by excessive subordination. We will classify this type of discordant postmodification as *subjunct[4] postmodification*. Now we can define the concordant and discordant varieties of hierarchic postmodification as follows:

Rule
concordant postmodification $\longrightarrow$

subjunct[3] postmodification

subjunct[1] postmodification

subjunct[0] postmodification

Rule
discordant postmodification $\longrightarrow$

subjunct[4] postmodification

Noun phrases. We will now combine premodification and postmodification to define a larger constituent, the *noun phrase*.[46]

Rule
noun phrase $\longrightarrow$

(premodification)*   noun   (postmodification)*

---

45 The meanings of *and* and *or* in our grammar are explained in Appendix A.
46 In the full grammar, a noun phrase can also be a pronoun or a nominal clause.

The following stylistic variations of noun phrases are defined in the obvious way, according to the corresponding types of their premodification and postmodification:

monoschematic noun phrase

centroschematic noun phrase

heteropoisal noun phrase

concordant noun phrase

discordant noun phrase

A question that arises at this point is the *projection problem;* that is, the problem of how the stylistic effect of a sentence, or even a sentence component, is modified by the stylistic characteristics of the components nested within it. Our solution is to assume that if a component is concordant then all its components, to all levels of nesting, must be concordant, but if a component is discordant then at least one of its components, at some level of nesting, must have been discordant. Thus, if we have a noun phrase consisting of concordant premodification, a noun, and discordant postmodification, the whole noun phrase will be classified as discordant.

We realize this is a very simplified solution, but we believe it will be acceptable in our work, for our sample corpus, although drawn from high-quality magazine writing, has few examples of nesting to more than one or two levels.

Verb phrases. The basic *verb phrase* is defined as follows:

Rule
verb phrase $\longrightarrow$

(adverbial)   verb   (adverbial)   (complement)*

The following stylistic variations of verb phrases are defined in the obvious way, according to the types of their components:

monoschematic verb phrase

centroschematic verb phrase

heteropoisal verb phrase

concordant verb phrase

discordant verb phrase

Sentences. We also define rules in the full grammar for complements, prepositional phrases, and dependent clauses. With these rules, we have all the components of a sentence and can define sentence structures of increasing complexity. The first of these structures is the *major,* a single main clause; the next is the *complete,* which adds dependent clauses to the main clause. In some rules, we introduce specializations according to the position of a certain type of stylistic element. An *initial discordant complete,* for example, is a sentence that has a discordant component in the initial position.

The basic major is defined as follows:

Rule

major $\longrightarrow$

(conjunction) (adjective)* (adverbial)* (prepositional phrase)* (nominal group)*
noun phrase   verb phrase

We define the following stylistic specializations of major sentences, according to the types of their components:

monoschematic major

centroschematic major

heteropoisal major

initial heteropoisal major

medial heteropoisal major

final heteropoisal major

concordant major

discordant major

Now we allow initial or final clauses to be added to the basic major sentence to give a *complete sentence*:

Rule

complete $\longrightarrow$

(clause)*   major   (clause)*

We define stylistic specializations of complete sentences in a manner analogous to those for major sentences, so that our grammar includes the following varieties:

monoschematic complete

centroschematic complete

heteropoisal complete

initial heteropoisal complete

medial heteropoisal complete

final heteropoisal complete

concordant complete

discordant complete

initial concordant complete

medial concordant complete

final concordant complete

initial discordant complete

medial discordant complete

final discordant complete

4.3 Grammar of Abstract Elements

Having completed the construction of the bottom level of the stylistic grammar, the classification of primitive stylistic elements, we can now define the central level, the grammar of abstract elements. This will correlate the stylistic effects of the primitive elements with the abstract elements that were defined in Section 3.3. At this level, the connective and hierarchic views become integrated.

A *monoschematic* sentence is a single main clause with optional, simple forms of subordination.

Rule
monoschematic ⟶

monoschematic complete

A *centroschematic* sentence is built up around a central main clause and can have complex subordination and dependent clauses.

Rule
centroschematic ⟶

concordant complete

A *polyschematic* sentence is built up around at least two central main clauses and has at least one dependent clause.

Rule
polyschematic ⟶

concordant complete   (concordant complete)$^+$

A *homopoisal* sentence is a coordination of syntactically similar structures. Here, we simplify to allow coordination of only very basic sentences.

Rule
homopoise ⟶

monoschematic complete   (monoschematic complete)$^+$

A *heteropoisal* sentence has at least one parenthetical component.

Rule
heteropoise ⟶

initial heteropoisal complete

medial heteropoisal complete

final heteropoisal complete

The various types of *concord* and *discord* are defined according to the presence of a concordant or discordant component in *initial, medial,* or *final* position.

Rule
  initial concord $\longrightarrow$

    initial concordant complete

Rule
  medial concord $\longrightarrow$

    medial concordant complete

Rule
  final concord $\longrightarrow$

    final concordant complete

Rule
  initial discord $\longrightarrow$

    initial discordant complete

Rule
  medial discord $\longrightarrow$

    medial discordant complete

Rule
  final discord $\longrightarrow$

    final discordant complete

As well as observing localized effects of concord and discord, we can recognize more-global shifts from concord to discord, or discord to concord. That is, a specific part of the sentence seems incongruous, discordant, but this discord is interpreted within the context of the rest of the sentence, so that it may either be 'resolved' by a return to a usual form, or left unresolved. We have formalized these two situations in the abstract elements of *resolution* and *dissolution*. A *resolution* is a shift in stylistic effect that occurs at the end of a sentence and is a move from a relative discord to a concord. A *dissolution* is a shift in stylistic effect that occurs at the end of a sentence and is a move from a relative concord to a discord.

Rule
  resolution $\longrightarrow$

    initial discord *and* final concord

Rule
dissolution $\longrightarrow$

    initial concord *and* final discord

4.4 Grammar of Stylistic Goals
Having completed the central level in the stylistic grammar, the grammar of abstract elements, we can now define the rules at the top level that correlate patterns of the abstract elements with the writer's specific goals.

    Stylistic goals, such as clarity, are elusive qualities that up to now have been defined by stylists by means of examples and informal rules. However, with the vocabulary and formal grammar that we have defined, we now have a way of seeing and abstracting what these examples have in common. We can abstract from a plethora of low-level syntactic rules that stylists traditionally have used, and can now define formal rules for specific stylistic goals.

    Note that stylistic goals can be organized along orthogonal dimensions. For example, a writer might try to be clear, or obscure, or make no effort either way. *Clarity* and *obscurity* are thus opposite ends of a stylistic dimension. Likewise, the goals of *concreteness* and *abstraction* form a dimension, and so do *staticness* and *dynamism*. Below, we look at one end of each of these dimensions; for details of their duals, see DiMarco (1990) and Hoyt (1993).

Clarity. *Clarity* is a very pervasive stylistic goal. Almost all the advice in textbooks of style and rhetoric is aimed at teaching the writer how to achieve clarity. Clarity, in other words, is thought to be the norm that a writer should strive for. To be clear is to be plain, precise, and predictable. In keeping with stylistic theory, therefore, we interpret the goal of clarity as adherence to the stylistic norm and incorporate accepted definitions of norm, that is, cohesive and concordant structures, in our grammar rule for clarity.

    Thus, we can use our stylistic grammar to give precise definitions to the kinds of sentences that Kane (1983), for example, associates with clarity:

- *Simple* sentences, which consist of one independent clause: these are our *monoschematic* sentences.

- *Centered* sentences, which consist of dependent constructions, followed by a main clause, followed by additional dependent clauses. We expand the notion of 'centered sentence' to include any sentence in which there is a dominant, concordant core: these are our *centroschematic* sentences.

- *Parallel* sentences, which reduce ambiguity by stressing the same grammatical form: these are our *homopoisal* sentences.

    Hence, we define clarity as follows:

Rule
clarity $\longrightarrow$

    monoschematic

    centroschematic

    homopoise

Concreteness. Kane (1983) associates *concreteness* with sentences that suggest an effect of immediacy, in which the writer has arranged elements to reflect the natural order of events or ideas, so that syntax mirrors events. In the terms of our grammar, concreteness is associated with sentences that emphasize a particular component, which may be highlighted either because it is *discordant* or because it is parenthetical, as in a *heteropoise*.

Rule
  concreteness $\longrightarrow$

      initial discord

      medial discord

      final discord

      dissolution

      heteropoise

Staticness. We associate *staticness* with sentences in which there is little opportunity for stylistic variation, that is, sentences that verge on being 'fixed forms.' These are the *monoschematic* sentences, which are standard and simple structures, or the more elaborate but strictly balanced structures, the *homopoisal* sentences.

Rule
  staticness $\longrightarrow$

      monoschematic

      homopoise

5. STYLISTIQUE: A Syntactic Stylistic Parser

So far, we have developed a vocabulary and methodology for constructing a stylistic grammar and we have shown how these tools have been applied to the construction of an English syntactic stylistic grammar. In DiMarco (1990), we applied the same vocabulary and methodology to develop a French syntactic stylistic grammar. However, these grammars provide only a theoretical foundation for a computational theory of stylistics. Now we describe the implementation of the stylistic grammars in a computational system. In this section, we will describe the organization of STYLISTIQUE, a syntactic stylistic parser that interprets the stylistic grammar. We will present a short example to illustrate STYLISTIQUE's method of analysis.

5.1 How STYLISTIQUE works
The English and French syntactic stylistic grammars are implemented in separate parsers that form the STYLISTIQUE system, a definite clause grammar that consists of

```
det(conjunct2_determiner,subjunct1_determiner,this,singular).
det(conjunct2_determiner,subjunct1_determiner,my,singular).
adj(conjunct1_adjective,subjunct1_adjective,large).
```

Figure 2
Sample lexical entries.

over 9,000 lines of QUINTUS PROLOG code. Each stylistic parser consists of the following three major modules:

- **Lexicon:** In addition to conventional information, the lexical entries are augmented by annotations indicating the connective and hierarchic primitive stylistic classifications associated with each word. Figure 2 shows several sample lexical entries.

- **Syntactic analysis:** The *parser* performs a conventional syntactic analysis, building a parse tree that identifies the grammatical structure of a sentence.[47]

- **Stylistic analysis:** The stylistic analysis proceeds in tandem with the syntactic parse. The *primitive-element analysis* assigns connective and hierarchic classifications to each sentence component as it is added to the parse tree. The *abstract-element analysis* builds higher-level patterns from these connective and hierarchic primitive elements. The *stylistic-goal analysis* correlates these patterns of abstract elements with the author's possible stylistic goals. It chooses one goal from each of the following three dimensions:

  - Clarity/neutrality/obscurity;
  - Concreteness/neutrality/abstraction;
  - Staticness/neutrality/dynamism.

The translation from the grammar rules that were shown in Section 4 to the implementation was straightforward. In most cases, there was an easy way of mapping between the theoretical rules and the PROLOG rules. For example, *heteropoisal postmodification* is described in the grammar as postmodification that is parenthetical and either conjunct or antijunct, such as a nominal group or a nonfinite clause. The grammar rules and corresponding PROLOG rule for one alternative in the definition of postmodification that is both heteropoisal and concordant is shown in Figure 3.[48]

Although STYLISTIQUE can usually run unaided, it does require punctuation in the input sentences to assist in disambiguation during parsing. The underlying stylistic grammar is highly complex, and so, lacking the aid of semantics, the stylistic parser can produce a number of syntactically grammatical, but not necessarily stylistically correct, parses for a given sentence. The standard PROLOG cut might have constrained backtracking and reduced the number of correct parses, but it seemed too crude and unpredictable to be easily applied to so complex and interdependent a system of rules

---

47 The English parser is based on a conventional parser written by Kem Luther and Rick MacLean at the University of Toronto, with supplementary rules suggested by Crystal and Davy (1969) and Quirk et al. (1985). The French parser is also based on Luther and MacLean's code, with supplementary rules adapted from Dubois and Dubois-Charlier (1970) and Galichet (1970).

48 The prefix "c" in the PROLOG rule indicates that the structure is concordant.

heteropoisal postmodification $\longrightarrow$

　conjunct[3] postmodification *with* parenthesis

conjunct[3] postmodification *with* parenthesis $\longrightarrow$

　nominal group

```
build_styl_postmod(
  StyleTree,
  styl(c_postmodification,hp_postmodification,
  StyleTree)) :-

  % Concordant and heteropoisal postmodification
  smember(c_nominal_group,StyleTree).
```

Figure 3
Sample grammar and PROLOG rules.

as STYLISTIQUE. As a simpler solution, we introduced explicit markers of punctuation that provide a form of partial disambiguation, as they do what the system should have done unaided but couldn't, because of the limitations mentioned above. STYLISTIQUE requires this prior disambiguation mostly for reasons of efficiency. For example, most parsers are unable to handle the conjunction *and* by any means other than trying all possible parses (Snarr 1984). The problem is analogous for *or* and for the comma (which can be considered a word). In our parser, the situation is complicated by the same conjunctions and the same punctuation markers playing different roles according to their level in the sentence structure. As a consequence, different conjunctions and punctuation markers are used at the sentence, clause, complement, and noun phrase level.

5.2 A Sample Stylistic Parse
A corpus of 75 sentences (52 English, 23 French) was used to test the parser. These sentences were chosen to demonstrate a substantial degree of stylistic variation. The following short example illustrates the kind of analysis that STYLISTIQUE produces for the sentence:

　45. True, posterity has been kind.

In subsequent sections, we will give an interpretation of the following parse tree for this sentence.[49]

```
Stylistic goals of this sentence:
[clarity,concreteness,neutral]

------------------------------------------------
Abstract stylistic elements (Connective view):
[[initial_and_medial_concord,initial_concord],
 [centroschematic,monoschematic],
 [initial_heteropoise,c_initial_heteropoise]]

------------------------------------------------
```

---

49 For brevity, we have omitted the hierarchical primitive-element parse.

```
Abstract stylistic elements (Hierarchic view):
[[],
 [centroschematic],
 [initial_heteropoise,c_initial_heteropoise]]
---------------------------------------------

Connective stylistic parse:

  c_sentence
    c_complete
    c_initial_heteropoisal_complete
      c_major
      c_heteropoisal_major
      c_initial_heteropoisal_major
          adjectival_phrase
            adjectival_phrase
              conjunct1_adjective
                 true
        c_noun_phrase
          c_noun_phrase
            c_nominal_group
              c_nominal_group
                c_premodification
                  c_premodification
                    c_premodification
                noun
                   posterity
                c_postmodification
                  c_postmodification
        c_verb_phrase
          c_verb_phrase
            xcopula
              copula
                been
            adjectival_phrase
              conjunct1_adjective
                kind
```

5.2.1 Primitive-Element Analysis.

The sentence is concordant, for it consists of a concordant main clause, the major, with no subordinate clauses. It begins with a style disjunct, *true*, which is an elliptic adjectival and therefore considered to have a connective, concordant effect, even if used in the initial parenthetical position. After the initial disjunct, the sentence continues with the bare noun *posterity*, which, lacking both premodification and postmodification[50] is a minimal, and therefore concordant, noun phrase. The sentence ends with the basic verb phrase *has been kind*, consisting of only the copula[51] *been*, and the concordant, *conjunct[1]* adjective *kind*; this is an inherently concordant verb phrase.

The sentence is concordant from the hierarchic view as well, for it has the form of a concordant initial heteropoisal complete sentence. This indicates that the sentence begins with a parenthetical construction, which in this case is the disjunct, *true*, a superordinate adjectival. The bare noun *posterity*, lacking both premodification and postmodification, is a monoschematic noun phrase. The verb phrase *has been kind* is basic and therefore monoschematic.

---

50 STYLISTIQUE must assign a primitive classification to the postmodification so that this information can be passed up to the higher levels of the noun phrase. As a consequence, the absence of postmodification must be marked as (trivially) centroschematic postmodification, in the connective view, and (trivially) monoschematic postmodification, in the hierarchic view.
51 The stylistic effects of auxiliary verbs are not taken into account, and so these verb forms are not recorded in the stylistic parse.

5.2.2 Abstract-Element Analysis. In the connective view, the significant position elements are initial and medial concords. The significant dominance elements are centroschematic and monoschematic. That is, this sentence has one dominant shape, which is, in fact, the whole sentence: it is monoschematic but also trivially centroschematic. STYLISTIQUE produces the most detailed analysis it can and does not prune extraneous information. The single important balance element is an initial heteropoise.

In the hierarchic view, STYLISTIQUE finds no significant position elements, but recognizes one dominance and one balance element: the sentence is centroschematic and an initial heteropoise. It is the initial disjunct, *true*, that introduces a superordinate effect; this feature makes the sentence slightly too complex to be monoschematic.

5.2.3 Stylistic-Goal Analysis. First, we consider the analysis of the sentence on the clarity/obscurity dimension. The presence of the concords in the connective view, together with the connective and hierarchic centroschematic structures, gives the sentence an effect of clarity. In a less obvious manner, the presence of an initial disjunct affects stylistic goals on the other dimensions. Because a superordinate parenthetical component is present, the sentence is a heteropoise and therefore considered to be concrete. On the staticness/dynamism dimension, the sentence has no definite leaning in either direction.

To summarize, this is a simple clear sentence with the slight incongruity of an initial parenthesis to relieve its blandness.

5.3 Limitations of the Implementation

STYLISTIQUE's limitations arise from the following characteristics of the system:

- The implementation of the grammar is incomplete.

- The grammar is too coarse-grained.

In the first case, the incomplete implementation of the grammar, STYLISTIQUE can sometimes produce analyses that are inaccurate, though not incorrect, because, due to time constraints, only about 90% of the full English and French syntactic stylistic grammars was implemented.

In the second case, the coarse grain of the stylistic grammar reflects theoretical limitations. The effect of coarse-grainedness was observed in the parses of some of the sentences in the sample corpus:

- Some sentences were parsed as simultaneously having both an initial *concord* and an initial *discord*. For example, both the following sentences were analyzed in this inconsistent manner:

    46. To tell everybody is the best thing.
    47. Telling lies is wrong.

- Some position elements were incorrectly identified: for example, a *final* discord was recorded as an *initial* discord.

- Some sentences with obviously different degrees of cohesiveness received the same analysis at the abstract-element level.

These anomalous results are not due to programming errors, but to the excessive ab-

stractness of the abstract elements and the resulting coarse grain of the grammar. It was a deliberate choice to define elements that were quite abstract. A major limitation of previous work in stylistics was the use of terms that were so specific and so numerous that it was not possible to identify general stylistic features that were common to sentences that were stylistically similar, but not obviously syntactically similar. However, our stylistic terms have occasionally erred on the side of being *too* abstract. For example, an *initial discord* in a STYLISTIQUE analysis may correctly identify the presence of a discord in the sentence but, because the scope of the element can be so broad, STYLISTIQUE sometimes cannot distinguish between a true initial discord, which occurs at the start of a sentence, and a noninitial discord that seems to be 'initial' because it affects the whole sentence.

The excessive abstractness of the stylistic elements was further demonstrated by the rules in Section 4.4, where we used the same patterns of elements to define both clarity and staticness. That there may be different types of monoschematic sentences, some merely clear, others static as well, has been overlooked in a stylistic analysis that relies on maximally expressive descriptions. Our grammar currently does not allow for such subtleties; although it fulfills our objective of a formalization of style that captures many generalities, it is not yet sufficiently expressive to distinguish all the subtleties we would wish.

6. Summary and Conclusion

In this final section, we will review the contributions of the research described in this paper and conclude with a discussion of some of the new directions that the work is taking.

6.1 Contributions of the Research

The problem of style presented advantages as a focus for new research. The codification of stylistic knowledge had been a virtually unexplored problem even within the general research area of computational linguistics. With very few exceptions, previous work had been unambitious (for example, counting word frequencies, or advocating basic rules of composition).

Our aim was to create a formal representation of goal-directed, nonliterary stylistics and, moreover, to do so in a manner applicable to different languages. The solution we proposed was the codification of stylistic knowledge in the form of a *stylistic grammar*. The construction of a stylistic grammar constitutes a theoretical advance over previous work in stylistics, for researchers had not attempted to produce a formal treatment of style, but had relied simply on unstructured normative or descriptive 'rules.' The work we have done toward a grammar of style has brought together ideas from stylistic theory and knowledge representation and applied them to a hitherto unformalized body of knowledge.

As a result of building English and French stylistic grammars, we were able to give more-formal definitions of stylistic goals. Previously, our understanding had been either purely subjective or based on established but unformalized usage. Now we have a grammar that correlates stylistic goals with specific patterns of abstract properties of text. Our contribution to more-formal definitions of stylistic goals was demonstrated for both syntactic and semantic style, as Ryan (1989, 1992) adapted our vocabulary and methodology to construct a semantic stylistic grammar that correlated the focus structure of paragraphs with the abstract elements, and the abstract elements with specific stylistic goals. In so doing, he augmented the definitions of stylistic goals.

The English and French syntactic stylistic grammars were implemented in STYLIS-

TIQUE, a stylistic parser that produced detailed goal-directed stylistic analyses of sentences typical of sophisticated magazine writing.

6.2 Applications of the Research
6.2.1 Style in Machine Translation.
We have developed separate English and French stylistic grammars and parsers. DiMarco and Hirst (1990) describes the application of our ideas to machine translation. Given the current French and English stylistic grammars, the next step in building a system that could preserve style in translation was to define a mapping between these grammars. Mah (1991, 1992) adapted and extended our work to add a practical, computational treatment of French–English comparative stylistics to the theory. This work will eventually make possible machine translation systems that would be able to preserve or modify style in translation. Makuta-Giluk (1991; Makuta-Giluk and DiMarco 1993) developed a computational theory of rhetoric that builds upon our theory to deal with the codification of higher-level pragmatic effects of language, such as formality, persuasion, and sincerity. Her work is also applicable to the preservation of stylistic effects in translation.

6.2.2 Natural Language Generation with Stylistic Constraints.
The ability to deal with stylistic and pragmatic aspects of language is important not only in natural language understanding, but in generation as well. Our stylistic grammar provides a formal representation of stylistic knowledge that was previously lacking in generation systems that attempted to deal with pragmatic issues. BenHassine (1992) adapted our knowledge representation for stylistics in order to incorporate stylistic constraints into the Penman language generation system (Penman 1988). Green (1992a, 1992b) extended and refined our theory of style by adapting work from functional grammar. Hoyt (1993) and Green have implemented the new theory in an integrated stylistic analyzer and generator, respectively. DiMarco, Hirst, and Stede (1993) are looking at lexical choice in natural language generation, including considerations of lexical style.

6.2.3 Second-Language Teaching.
What has been learned from developing STYLIS-TIQUE has also been applied to machine-aided language instruction. Existing language-teaching systems focus almost exclusively on the basics of composition. An instructional version of STYLISTIQUE could systematically develop a student's understanding of the more advanced aspects of language composition. Payette (1990; Payette and Hirst 1992) developed an instructional system that, applying some of our ideas, analyzes input sentences for basic normative style and clarity, and offers feedback to the student.

6.3 Conclusion
Stylistic and pragmatic aspects, though necessary in complete understanding of language, have been neglected in computational linguistics research. These problems had been too vague and ill-defined to be dealt with by computational systems. However, in this work, we have developed a novel, formal representation of stylistic knowledge that makes the problem of stylistic analysis more amenable to computational solution.

It is hoped that this research will lead to a system sophisticated enough to deal with a range of stylistic problems. Long-term applications include the development of a stylistic post-editor for use in a machine translation system. In addition, the continuing enhancements of the stylistic analyzer should contribute to a better understanding of the role style plays in language generation and teaching. The ongoing development of a formal framework for the representation of knowledge about stylistics should provide a partial computer model of how people produce style in language.

References

BenHassine, Nadia (1992). *A formal approach to controlling style in generation.* Master's thesis, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada.

Biber, Douglas (1988). *Variation across Speech and Writing.* Cambridge University Press.

Biber, Douglas (1989). "A typology of English texts." *Linguistics,* 27, 3–43.

Bureau, Conrad (1976). *Linguistique fonctionnelle et stylistique objective.* Presses Universitaires de France.

The Chicago Manual of Style. (1952). University of Chicago Press.

Cluett, Robert (1976). *Prose Style and Critical Reading.* Teachers College Press, Columbia University.

Cluett, Robert (1990). *Canadian literary prose: A preliminary stylistic atlas.* ECW Press.

Corbett, Edward P. J. (1971). *Classical Rhetoric for the Modern Student.* Oxford University Press.

Crystal, David, and Davy, Derek (1969). *Investigating English Style.* Longman Group Ltd.

DiMarco, Chrysanne (1990). *Computational stylistics for natural language translation.* Doctoral dissertation, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. Published as

technical report CSRI-239.

DiMarco, Chrysanne, and Hirst, Graeme (1990). "Accounting for style in machine translation." In *Proceedings, Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages.* Austin, Texas, June 1990, 65–73.

DiMarco, Chrysanne; Green, Stephen J.; Hirst, Graeme; Mah, Keith; Makuta-Giluk, Marzena; and Ryan, Mark (1992). "Four papers on computational stylistics." Research report CS-92-35, Faculty of Mathematics, University of Waterloo, June 1992.

DiMarco, Chrysanne; Hirst, Graeme; and Stede, Manfred (1993). "The semantic and stylistic differentiation of synonyms and near-synonyms." *AAAI Spring Symposium on Building Lexicons for Machine Translation,* Stanford, CA, March 1993, 114–121.

Dubois, Jean (1969). *Grammaire structurale du français: La phrase et les transformations.* Librairie Larousse.

Dubois, Jean, and Dubois-Charlier, Françoise (1970). *Éléments de linguistique française: syntaxe.* Librairie Larousse.

Fowler, H. W. (1968). *A Dictionary of Modern English Usage.* Second edition. Oxford University Press.

Galichet, Georges (1970). *Grammaire structurale du français moderne.* Éditions HMH.

Green, Stephen (1992a). "A basis for a formalization of linguistic style." In *Proceedings, 30th Annual Meeting of the Association for Computational Linguistics.* Newark, Delaware, June 1992, 312–314.

Green, Stephen (1992b). *A functional theory of style for natural language generation.* Master's thesis, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada. Published as research report CS-92-48, Faculty of Mathematics, University of Waterloo.

Grevisse, Maurice (1986). *Le bon usage.* Duculot.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar.* Edward Arnold.

Halliday, M. A. K., and Hasan, Ruqaiya (1976). *Cohesion in English.* Longman Group Ltd.

Hendricks, William O. (1976). *Grammars of Style and Styles of Grammar.* North-Holland.

Heritage, John (1984). *Garfinkel and Ethnomethodology.* Polity Press.

Hovy, Eduard (1988). *Generating Natural Language under Pragmatic Constraints.* Lawrence Erlbaum Associates.

Hoyt, Pat (1993). *A functional-based stylistic analyzer*. Master's thesis, Department of Computer Science, University of Waterloo, October 1993. Published as research report, Faculty of Mathematics, University of Waterloo.

Jameson, Anthony (1987). "How to appear to be conforming to the 'maxims' even if you prefer to violate them." In *Natural Language Generation*, edited by Gerard Kempen. Martinus Nijhoff.

Jensen, Karen; Heidorn, George; Richardson, Stephen; and Haas, Norman (1986). "PLNLP, PEG, and CRITIQUE: Three contributions to computing in the humanities." *Computers and the Humanities Conference Guide*, Toronto, Ontario, April 1986, 183–193.

Kane, Thomas S. (1983). *The Oxford Guide to Writing*. Oxford University Press.

Lanham, Richard (1974). *Style: An Anti-Textbook*. Yale University Press.

Loffler-Laurian, Anne Marie (1986). "Stylistics and computing: Machine translation as a tool for a new approach to stylistics." *Computers and Translation*, 1, 215–223.

Mah, Keith (1991). *Comparative stylistics in an integrated machine translation system*. Master's thesis, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada. Published as research report CS-91-67, Faculty of Mathematics, University of Waterloo.

Mah, Keith (1992). "A computational formalism for French–English comparative stylistics." In DiMarco et al. (1992), 8–13.

Makuta-Giluk, Marzena (1991). *A computational rhetoric for syntactic aspects of text*. Master's thesis, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada. Published as research report, CS-91-56, Faculty of Mathematics, University of Waterloo.

Makuta-Giluk, Marzena and DiMarco, Chrysanne (1993). "A computational formalism for syntactic aspects of rhetoric." *Proceedings, The First Conference of the Pacific Association for Computational Linguistics*, Vancouver, Canada, April 1993, 63–72.

Milic, Louis T. (1982). "The annals of computing: Stylistics." *Computers and the Humanities*, 16(1), September 1982, 19–24.

Nunberg, Geoffrey (1990). *The Linguistics of Punctuation*. CSLI Lecture Notes 18.

Center for the Study of Language and Information, Stanford University, Stanford, CA.

Payette, Julie (1990). *Intelligent computer-assisted instruction in syntactic style*. Master's thesis, Department of Electrical Engineering, University of Toronto, Toronto, Ontario, Canada. Published as technical report CSRI-247.

Payette, Julie, and Hirst, Graeme (1992). "An intelligent computer-assistant for stylistic instruction." *Computers and the Humanities*, 26, 87–102.

Penman documentation. (1988). The Penman Natural Language Generation Group, Information Sciences Institute, University of Southern California.

Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; and Svartvik, Jan (1985). *A Comprehensive Grammar of the English Language*. Longman Group Ltd.

Raskin, Robin (1986). "The quest for style." *IEEE Transactions on Professional Communication*, PC-29(3), September 1986, 10–18.

Ryan, Mark B. P. (1989). *The computational codification of the semantic aspects of style*. Master's thesis, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. Published as technical report CSRI-231.

Ryan, Mark B. P. (1992). "Focus shifts as indicators of style." In DiMarco et al. (1992), 1–7.

Scott, Donia R., and de Souza, Clarisse Sieckenius (1990). "Getting the message across in RST-based text generation." In *Current Research in Natural Language Generation*, edited by Robert Dale, Chris Mellish, and Michael Zock. Academic Press.

Snarr, Victoria L. (1984). *Theory and parsing of the coordinate conjunction "and."* Master's thesis, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. Published as technical report CSRI-171.

Strunk, William Jr., and White, E. B. (1979). *The Elements of Style*. Third edition. Macmillan.

Tsutsumi, Taijiro (1990). "Wide-range restructuring of intermediate representations in machine translation." *Computational Linguistics*, 16(2), June 1990, 71–78.

Vinay, J.-P., and Darbelnet, J. (1958). *Stylistique comparée du français et de l'anglais*. Didier.

Appendix A: Notes on Terminology

At all levels of the grammar, the left-hand side of each rule identifies what is being defined, and the right-hand side lists one or more alternative realizations, one per line. Sometimes, a particular alternative will be illustrated by an accompanying example. For instance, the following rule describes the *conjunct²* type of *adjectival*:

Rule
  conjunct² adjectival ⟶

      premodifying genitive
        his *religious works*

      demonstrative determiner
        this *substantial selection*

In this rule, the first alternative realization of a *conjunct² adjectival* is a *premodifying genitive*, such as the possessive *his* in *his religious works*. The second alternative is a demonstrative determiner, such as *this* in *this substantial selection*.

In the grammar, we will use various shorthand notations to simplify the presentation of the rules. However, these abbreviated forms can be expanded into standard context-free grammar rules. The shorthand notations are as follows; they are illustrated by particular examples, but are intended for general use:

1.   adjectival ⟶ intensifier   adjective

The juxtaposition of terms on the right-hand side of a rule indicates a concatenation of instances of these terms. For example, the rule above allows the intensifier *very* to be followed by the adjective *happy* to form an adjectival, *very happy*.

2.   adjectival ⟶ (intensifier)   adjective

Parentheses indicate that the form is optional. In this example, an adjectival could be either an intensifier followed by an adjective or an adjective alone.

3.   adjectival ⟶ (intensifier)⁺   adjective

The Kleene cross indicates one or more occurrences of the form within parentheses.

4.   adjectival ⟶ (intensifier)*   adjective

The Kleene star indicates zero or more occurrences of the form within parentheses.

5.   postmodification *with* parenthesis

Where a rule has several alternatives, this shorthand notation using *with* abbreviates a long sequence of alternatives (here, the many types of parenthesis).

6.   concordant heteropoisal postmodification ⟶
          concordant postmodification *and* heteropoisal postmodification

And indicates that all conditions on the right-hand side of a rule must simultaneously be satisfied by a single constituent.

> 7.   concordant heteropoisal postmodification $\longrightarrow$
>   nominal group *or* prepositional phrase

Or indicates that any one of the conditions on the right-hand side of a rule must be satisfied.